THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XXXV NUMBER 2 JUNE 1989

PUBLISHED BY
THE UNIVERSITY OF ALBERTA • EDMONTON

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

Published in March, June, September, and December
by the Faculty of Education, University of Alberta

A quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

EDITOR: Thomas O. Maguire

CONSULTING EDITORS

Reno A. Bosetti
Alberta Department of Education

Robert C. Bryce
University of Regina

Janet A. Donald
McGill University

Michael G. Fullan
University of Toronto

Jane Gaskell
University of British Columbia

Myrna L. Greene
University of Lethbridge

Lilian G. Katz
University of Illinois

Joanne C. Keselman
University of Manitoba

John McLeod
University of Saskatchewan

Alan R. Pence
University of Victoria

Lauran H. Sandals
University of Calgary

Hugh A. Stevenson
University of Western Ontario

Keith C. Sullivan
Dalhousie University

Jaap Tuinman
Simon Fraser University

John D. Wilson
Moray House College of Education

FACULTY ADVISORY COMMITTEE

Harvey Zingle (Chair)
Alton T. Olson
Walter H. Worth
Daiyo Sawada

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada and from the Alberta Advisory Committee for Educational Studies.

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to *The Alberta Journal of Educational Research*, Faculty of Education, Publication Services, 4-116 Education North, University of Alberta, Edmonton, Alberta, Canada, T6G 2G5.

SECOND CLASS MAIL REGISTRATION NUMBER 1436

Faculty of Education
University of Alberta

ISSN 0002-4805

The Alberta Journal of Educational Research

Volume XXXV, Number 2 *June 1989*

CONTENTS

THE EDITOR'S PAGE: Educational Research—
Indigenous or Exotic? 111

M.E. LaZerte: Pioneer Educational Innovator 112
G.H. Buck

Evaluation of Student Achievement:
Teacher Practices and Educational Measurement 123
J.O. Anderson

Evaluating Student Achievement
in an Ontario High School 134
R.J. Wilson

Word Processing Effects on Writing
Productivity and Revision at Elementary
and Junior High School Levels 145
L.L. Morton, P.H. Lindsay, and W.M. Roche

Levels of Performance and Cognitive
Interference in Test-Anxious Subjects 164
C.S. Hammermaster

The College to University Experience:
Satisfaction and Success of
Students Who Transfer 171
L.D. Vaala and E.A. Holdaway

The Alberta Journal of Educational Research *Vol. XXXV, No. 2, June 1989, 111-111*

THE EDITOR'S PAGE

Educational Research—Indigenous or Exotic?

In early December 1988, I attended the annual conference of the Australian Association for Research in Education. The theme of the conference was the title for this editorial. Like us, the Australian educational research community is strongly influenced by American ideas. Many of the researchers were trained in the United States; the journals of the American educational research community are viewed as being more desirable outlets than national journals; and many of the dominant research themes are those that are on the American agenda. The question that was raised in a few of the sessions was what, if anything, should be done about it?

According to Richard White of Monash University, the influence is unidirectional. After examining many of the American research journals, White concluded that American authors seldom cite foreign published material, and when they do it is almost always a book rather than a journal. White correctly notes that the best chance that a researcher has of having an international influence on educational research is to publish in American or British journals. But what he doesn't answer is, "What effect does this have on the development of an indigenous approach to educational research?" Is there a case to be made for the development of a unique perspective on problems? I think that there is.

In my own field of educational measurement, many of the topics that appear in the American literature are driven by the social-legal context of American education. For example, minimum competency testing, certification testing for teachers, and item bias are all topics of major interest that have their roots in American legislation. There are examples from other fields—mainstreaming in special education comes immediately to mind. While it would be inaccurate to claim that all research published in American educational research journals must fit their national framework, the possibility of having a large audience for your work is a powerful reinforcer, and I suspect that some of our colleagues are influenced by it, not only in what they research, but by how they research it.

It is important that we encourage research from our own perspective, not only research on problems for which there is a Canadian concern, such as effects of French immersion, but also research on the *generic* problems. It may seem self-serving to ask you to make *AJER* your outlet of first choice, but examine the articles in this issue and see if you don't agree that they not only make a strong contribution to educational thought, but they also have a distinctive approach to them. To underline the point I have placed George Buck's paper on M.E. LaZerte as the lead article for this issue. LaZerte was a man who exemplified the kind of researcher that our community needs.

T.O.M.

The Alberta Journal of Educational Research *Vol. XXXV, No. 2, June 1989, 112-122*

GEORGE H. BUCK
University of Alberta

M.E. LaZerte: Pioneer Educational Innovator

M.E. LaZerte was a prominent educator in Alberta during the first half of this century. This article describes LaZerte's learning theories and several instructional devices he created and tested. Their merits and weaknesses are discussed from LaZerte's viewpoint, as well as their relationship to general instructional theory and pedagogy. The article also compares LaZerte's innovations and learning theories to other contributors, including those of the prominent American psychologist B.F. Skinner. LaZerte's findings from his instructional devices may have relevance in assessing the efficacy of microcomputers used for instruction.

It has been over 35 years since Milton Ezra LaZerte (1885-1975) held an official position in the University of Alberta. Why should his name be remembered and why should anyone still consider the merits and the findings of his research? Although LaZerte began his career in education in rural Alberta, he did not conform to the common yet erroneous notion that most rural teachers, especially those in Western Canada, possessed pedagogical skills inferior to those of teachers in the large centers of central and eastern Canada and the United States. He not only excelled as a teacher, he also demonstrated his skills in both educational innovation and administration early in his career (Chalmers, 1978).

LaZerte was the first director of the School of Education at the University of Alberta during the 1930s, and he later became the first Dean of that university's Faculty of Education; this was the first faculty in any provincial Canadian university that prepared all teachers for that province (Johns, 1981, pp. 192-193). Immediately following World War I and well before obtaining responsible administrative positions, LaZerte addressed the problems encountered by many teachers attempting to teach difficult concepts to large and often ungraded classes. A believer in scientific research methods as well as innovative experimentation, LaZerte realized that many of the problems experienced by teachers as well as students might be alleviated in part by instructional materials designed to be used by an individual student without constant supervision (LaZerte, 1922, p. 30). It is significant to note that LaZerte began his work in this area several years before the so-called "progressive" movement took root in North American education (Charyk, 1974, pp. 277-278; LaZerte, 1945, p. 12). As early as 1922, LaZerte noted a need for such instructional materials:

George Buck is currently working toward the completion of a doctoral program in educational psychology at the University of Alberta.

> There is little apparatus in our schools to assist the teacher in giving a practical setting to the work that covers the field from group counting in primary years to the development of the angle-sum theorem in grade VIII geometry. (p. 30)

During the past 30 years, many individuals and groups from various disciplines, particularly outside education, have attempted either to "improve" or to "save" North American education by means of various teaching and testing machines. Certain American psychologists made the claim in the mid-1950s that no educators had even considered, much less experimented with, automated teaching (Skinner, 1954). Clearly they were unaware of the work and publications of LaZerte. The behaviorist B.F. Skinner developed what he called "teaching machines" that he felt would release individual teachers from boring and onerous tasks and enable them to spend their time doing more important duties. Skinner also claimed that his devices produced a significantly higher level of achievement among his students at Harvard University (Skinner, 1983, pp. 139-140). He could not understand most educators' reluctance to commit themselves to the use of either his machines or other teaching machines (Skinner, 1984). Another psychologist, Sidney L. Pressey, claimed that he had invented the first teaching machines during the early 1920s and that no interest was shown in them by educators, even though, as he argued, they provided "great savings in material and labor" (Pressey, 1932, pp. 671-672).

For a brief period during the 1960s, interest in teaching machines seemed to wax among many educators. Consequently, several companies designed and produced a variety of devices (Hendershot, 1964). In practice, however, teaching machines did not achieve the goals which Skinner had imagined, nor did such machines see widespread use, especially in Canada (Rutherford, 1961). No one at that time seemed to know that M.E. LaZerte had experimented with and rejected the use of auto-instructional devices of his own design, using primary and secondary school children as his experimental subjects, at least 30 years before Skinner came on the scene to promote an interest in teaching machines (Charyk, 1974, pp. 277-278; LaZerte, 1933, p. 5). If LaZerte actually tried out teaching devices, what were his systems and the machines themselves like? Why was no attempt made to have them mass-produced and used in the schools in order to improve pedagogy and student learning? And how are they relevant to our age of microcomputers? The answers to these questions can be found by studying LaZerte's experiments and the machines themselves, by considering his findings, and by taking into account the mood and educational environment of the time.

Background

LaZerte received a B.Educ. degree (a graduate degree of the time, similar to the contemporary Master of Education degree) from the University of Alberta in 1926. His thesis, *A Study of the Methods Used by Elementary School Pupils in Solving Problems in Arithmetic*, contains some interesting conclusions related to learning theory. LaZerte (1926) stated that, "the ability of pupils to find correct answers to problems in arithmetic is a poor measure of their actual problem-solving ability" (p. 102). He also found that

"much problem-solving is but a slavish adherence to type methods of procedure which become associated with certain language and number situations that exist in the presented problem" (LaZerte, 1926, p. 102). He took the findings as indications that the pedagogical methods employed by many teachers were unsatisfactory because the action of ascertaining a correct answer to a mathematical problem does not indicate to the teacher whether the underlying problem-solving strategy or concept has been either learned or understood by the student. These findings were, of course, not new. The Roman orator and educator, Marcus Fabius Quintilianus, known as Quintilian, had made the same observations as early as the first century A.D. (Quintilian, 1. 1. 24-25). It is also possible that similar observations were made by others prior to Quintilian. The common findings of Quintilian and LaZerte indicate that many educators are conservative; they believe that "traditional" methods, no matter how demonstrably inappropriate and inferior, are not to be displaced by newer ones. Both LaZerte and Quintilian also concluded that repetitive instruction such as drill and practice and plain memorization do not provide the student with an adequate facility for problem solving. This conclusion is still valid, especially in the criticism of modern "task analyses" (LaZerte, 1933, pp. 3-4; Quintilian, 10. 2. 2-5). It is of interest to note that Quintilian's and LaZerte's beliefs are diametrically opposed to those of Pressey and Skinner. Both of them advocated a behavioristic approach to learning, one in which the notion of the individual student formulating concepts about the subject matter being taught is dismissed as nonsense (Skinner, 1954). Although LaZerte acknowledged that his findings were based on a limited sample, he intended to investigate the matter further (LaZerte, 1926).

LaZerte worked on his Ph.D. at the University of Chicago under Charles H. Judd, a contemporary of Edward L. Thorndike. LaZerte's dissertation *A Diagnosis of Difficulties Encountered in Solving Problems in Arithmetic*, which built on his work for his B.Educ. thesis, was completed in 1927. Judd, unlike the behaviorists, shared LaZerte's belief that it was necessary for a student to develop a concept of what was being taught, rather than simply being able to parrot it. Judd stated that, "while formal drill gives a superficial mastery of arithmetic processes, it fails utterly to give pupils that higher training in reasoning which might be derived from an effort to understand the meaning of number operations" (LaZerte, 1933, p. x). LaZerte endeavored to devise methods and apparatus, not only to evaluate students' understanding of problem-solving processes, but also to develop devices that could instruct logical problem-solving procedures so that a student might develop an understanding of the underlying concepts.

Early Innovations

Envelope Test

One of LaZerte's first methods for assessing pupils' knowledge of problem-solving procedures was the Envelope Test (LaZerte, 1933, p. 12). The subject was presented with a large envelope on which was typed a question. Two smaller envelopes, located inside, each had a different problem-solving

procedure typed on the outside. The subject was to select only one of the envelopes, the one that was believed to display the correct procedure. This in turn contained two more envelopes with further steps. After further selection, the student would be left with one envelope, inside of which was a card. The card contained the last step of the problem on one side and a message on the other. Only one of the cards in the envelope array stated "This is the correct card"; the others stated, "You are wrong. Go back to the beginning and try again" (LaZerte, 1933, pp. 12-13). Figure 1 illustrates the hierarchical nature of the envelope test.

LaZerte noted that the number of both cards and envelopes was altered to accommodate questions of different complexity. In addition, he stated that some sequences possessed intermediate blank cards or envelopes which indicated that the examinee was on the wrong track. This modification was intended to save time, serving to inform the student of an erroneous choice early in the procedure. Although a blank card or envelope informed the examinee that an error had been made, it did not provide any corrective hint or cue (LaZerte, 1933, p. 13). In this manner, the Envelope Test did not instruct the examinee, and the procedure tested only what the examinee knew at the time of the test. The Envelope Test possessed the capability, nevertheless, of providing instruction in problem-solving strategy, if it was desired. LaZerte had created a rudimentary branching program anticipating those found in many "scrambled books" or programmed texts produced 25 years later (Lumsdaine & Glaser, 1960).

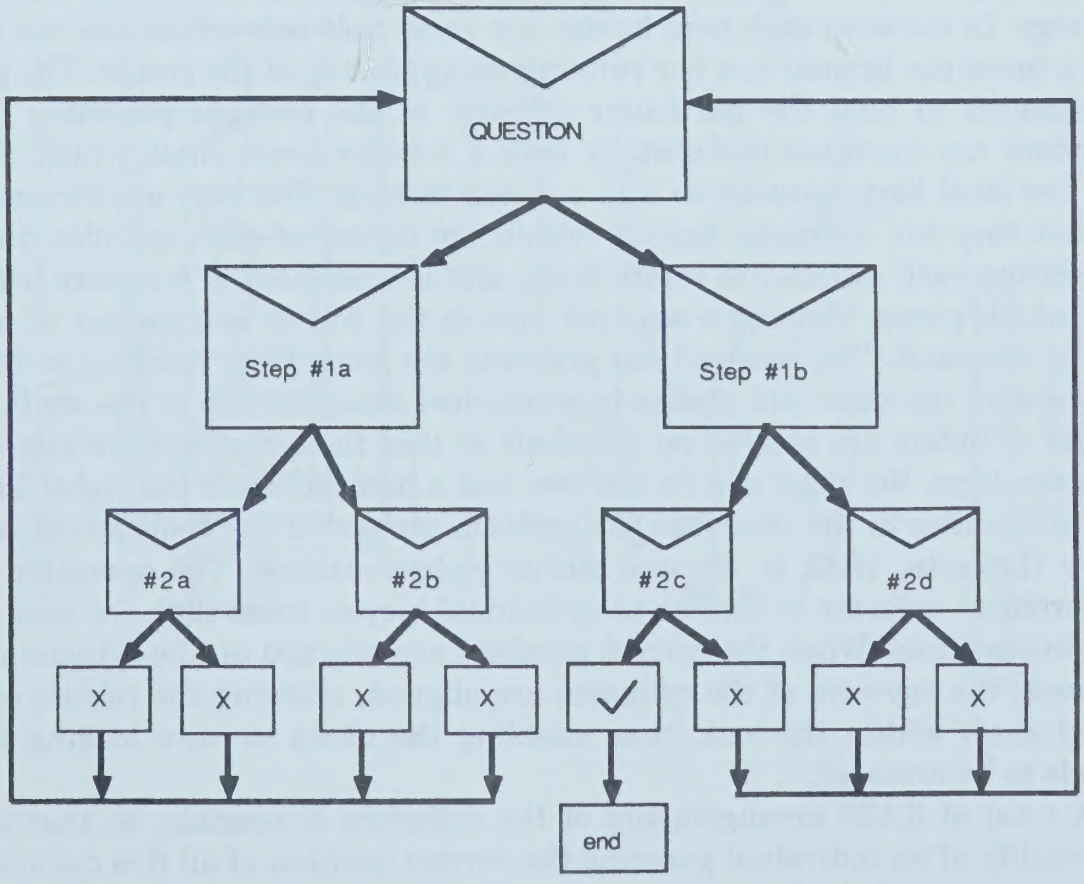


Figure 1: The hierarchical envelope test devised by LaZerte

The Problem Cylinder

There were distinct disadvantages to the Envelope Test. In order to prevent cheating, a supervisor had to be present while the test was in progress. The supervisor, as well, had to restore the cards and the envelopes after each test. Each Envelope Test, therefore, required the undivided attention of a supervisor who, if a different system were used, could have been supervising several examinations simultaneously. In addition, the relatively small number of decisions available to the examinee meant that one could possibly guess the correct sequence of envelopes as well as the correct card (LaZerte, 1933, p. 58). To alleviate these problems, LaZerte devised, and had constructed in 1930, a machine which he called the "problem cylinder" (Figure 2). Although others such as the psychologist Pressey were also developing testing machines at this time, there is no evidence in the literature or elsewhere to indicate that LaZerte's work was anything but an independent endeavor (personal communications with H. T. Coutts, 1987). There is no similarity in the concepts behind or the design of the machines or the learning theory implicit in its structure.

LaZerte's apparatus consisted of a wood and brass frame which housed five laminated wooden cylinders. The cylinders were constructed so that a keyway was present throughout their length. In addition, the center laminations were designed so that the cylinder could be rotated around a key if it was positioned in the center of the cylinder. Each cylinder possesses five brass card holders, evenly spaced about the circumference. The card holders are intended to hold segments of equations or portions of a problem-solving strategy. In between each card holder is a small hole into which one can insert a brass pin located in a bar running along the top of the frame. The pin is intended to hold the particular cylinder in the selected position. The cylinders are mounted horizontally onto a tubular brass shaft which also has five steel keys fastened to it in a linear fashion. The keys are arranged so that they are normally located within the center of each cylinder, thus permitting each cylinder to rotate freely and independently. A square brass rod, which passes through a squared hole in the frame, is attached to one end of the shaft. The squared rod prevents the shaft from rotating so that the keys on the shaft will always be positioned along the top of the shaft. If all the cylinders are rotated on the shaft so that their keyways are aligned with the keys, the shaft can be slid one and a half inches to the right. This action signifies to the user that the problem presented has been solved correctly (LaZerte, 1933, p. 59, and author's observations). The operation of the problem cylinder is similar to cylindrical bicycle locks that are used at the present time. When the correct numbers are selected on the cylinders of the lock, the keyways of the cylinders are aligned, allowing the release of a keyed shaft within the lock, thus enabling the chain or wire locking the bicycle to be removed.

A total of 3,125 arrangements of the cylinders is possible, so that the probability of an individual guessing the correct position of all five cylinders is extremely unlikely. In operation, the subject was presented with a mathematical problem and was then instructed to solve it, starting with the left-

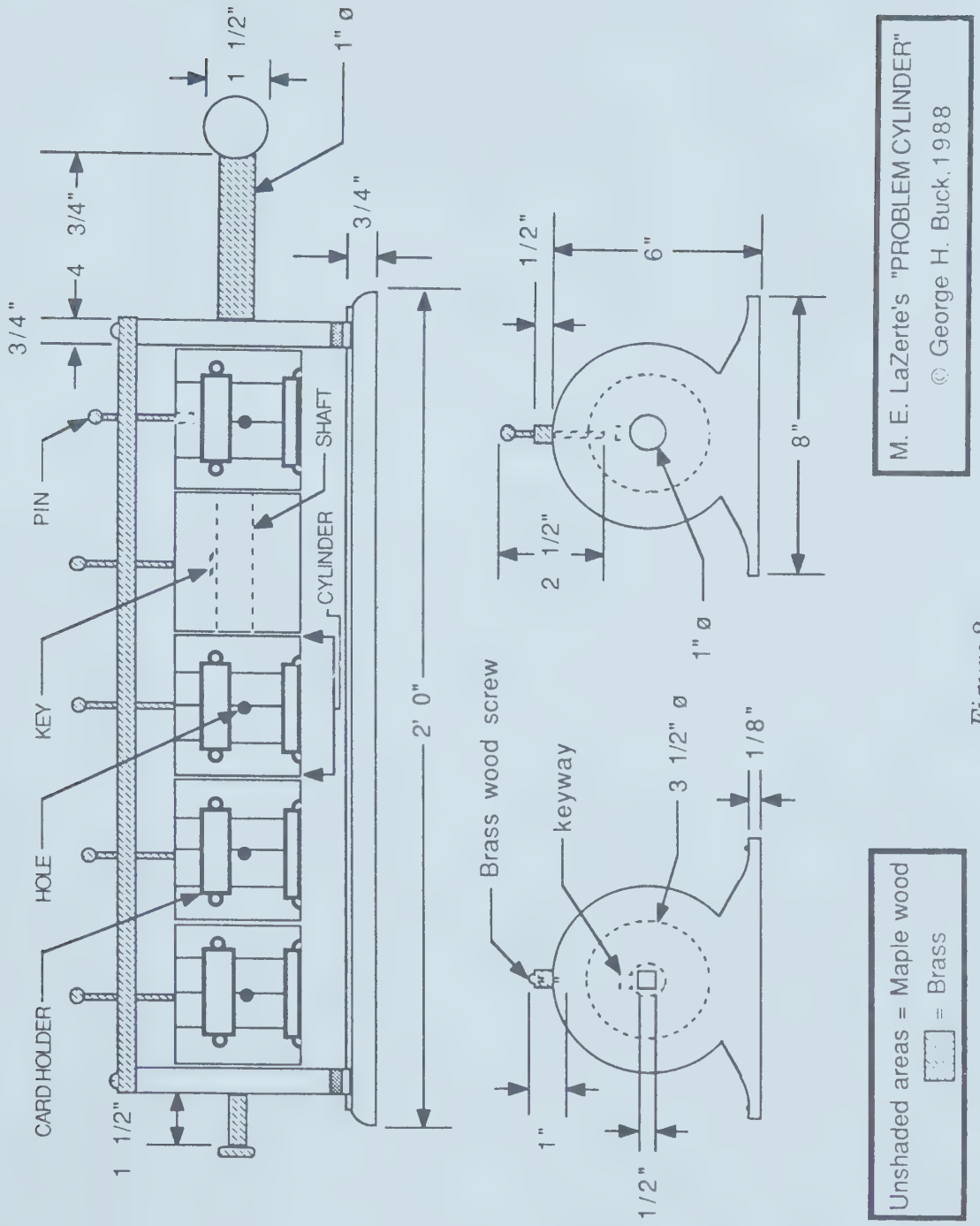

CARD HOLDER
HOLE
KEY
PIN
SHAFT
CYLINDER
1 1/2"
3/4"
4
3/4"
1 1/2"
1" ø
3/4"
2' 0"
Brass wood screw
keyway
1"
1/2"
3 1/2" ø
1/8"
2 1/2"
1/2"
6"
8"
1" ø
Unshaded areas = Maple wood
= Brass
M. E. LaZerte's "PROBLEM CYLINDER"
© George H. Buck, 1988

Figure 2

most cylinder and proceeding to the right. Every cylinder held five cards, each of which displayed a part of a problem solving strategy. The subject was to select what was believed to be the correct strategy among the five cards on each cylinder. This was accomplished by the subject rotating each cylinder so that the desired information faced him or her. After each selection was made, the subject was supposed to insert a brass pin into a hole in the top of the cylinder to prevent it from being moved while the next selection was undertaken. When it was felt that all five cylinders were in the correct position, the subject was to pull on the shaft. If the shaft did not slide to the right, one or more cylinders were improperly positioned, and this was an indication to the subject that a mistake in the steps of solving the problem had been made and that he or she should try again.

LaZerte (1933) found that the problem cylinder was intriguing to students and was highly motivational. He stated that, "all subjects were very much interested in the problem presented to them. After the experiment was completed many requested to continue their efforts. Motivation was strong throughout" (p. 60). In a fashion similar to that of the Envelope Test, the problem cylinder could be arranged so that fewer than five cylinders could be used. In such cases, those cylinders not in use would not have cards inserted in their card holders. In addition, the cylinders not in use would be pinned in the unlocked position.

Unlike the Envelope Test, however, the problem cylinder could be used without constant supervision because there was no easy way in which the subject could cheat or sabotage the machine. Besides being a testing device, the problem cylinder could be used to demonstrate logic or the logical analysis of a problem. The question noted at the outset of this article comes to mind: Why was no attempt made to have the problem cylinder mass-produced and used in the schools in order to improve pedagogy and student learning? The answer is provided by LaZerte's findings. Although high motivation and interest were attributed to the problem cylinder in a kind of "halo" effect, LaZerte (1933) also discovered that the skill required to understand and to manipulate the apparatus added to the difficulty of the problem being solved: "It is apparent that the task of solving the problems on the cylinder is much more difficult than when presented in the usual form" (p. 65). Data gathered from comparison groups verified his findings (LaZerte, 1933, pp. 63-67). It seems probable, therefore, that if LaZerte's problem cylinder were to be used successfully as a "teaching machine," instructions and practice exercises would have to be provided so that the manipulation of the machine would not add to the difficulty of the problem presented. Thus much time and effort would have to be expended on training in the use of the machine. Mutatis mutandis, the same criticism may be applied to microcomputers used for instruction. In addition to being familiar with the physical hardware such as the keyboard and peripheral devices such as a mouse, the computer user must also know the peculiarities of the programs used. It is unlikely, therefore, that the user will be able to give full and undivided attention to the primary task until some time is spent in learning about the machine and the program being used. This phenomenon

has been experienced by most microcomputer users, including some who describe the frustration caused by the problem (Vargas, 1986). Given this drawback, LaZerte concluded that the use of a teacher was more efficient and more effective than the use of his problem cylinder.

While the problem cylinder might appear to be primitive by current technological standards, it must be kept in mind that in 1930 the areas of instructional devices and electronics were both extremely limited in their development. In addition, most small school boards then in existence possessed limited funds and no electricity; it therefore seems unlikely that technologically complex and expensive apparatus would or could have been purchased by these school boards. Had his experimental studies shown the problem cylinder to be a successful teaching device, or if LaZerte, like Skinner 25 years later, had decided to market the device despite indications that it would not be successful in the education market, the problem cylinder could probably have been manufactured inexpensively because of the small number of parts. If the brass parts were replaced by steel or aluminum, and the wooden parts replaced with Bakelite, the cost could have been reduced even further.

Later Innovations

Activity Oriented Mathematics

Between the mid-1930s and the mid-1940s, LaZerte's time was primarily consumed by administrative obligations, but his interest in developing testing and instructional devices did not disappear. By the end of World War II, LaZerte had been testing new "activity" mathematics programs in the University of Alberta's demonstration school (LaZerte, 1945, p. 12). The programs, intended for primary grades, entailed considerable laboratory time for each pupil. LaZerte (1945) noted that, "the pupils spent 60% of the time devoted to arithmetic in the laboratory" (p. 12). In addition, the program was intended to be primarily individualized instruction where the students learned largely on their own and the teacher supervised. In this case, "individualized instruction" did not refer to the following of a rote program individually. It was expected that each student would perform the experiments outlined and discover for himself or herself the concept or concepts involved (LaZerte, 1951, p. 1). The laboratory work consisted of the use of instructional aids that were intended to facilitate the concrete representation of concepts, as well as to facilitate problem solving. The apparatus included bead frames, various number boards and charts, special colored disks, cut-outs, and rudimentary measuring devices (LaZerte, 1945, p. 12). While it produced favorable results, LaZerte noted that this "activity program" was similar to the methods he had used prior to 1930, "before the progressive movement reached Alberta" (LaZerte, 1945, p. 12). This work culminated in the production of the combination text and activity manual entitled *Numbers Tell Their Story*, first published in 1953.

Although the problem cylinder had been considered unsatisfactory for individualized instruction, LaZerte continued to believe that instructional apparatus could be useful in certain situations. LaZerte (1953) stated that,

"the emphasis on the use of apparatus and equipment should prove of inestimable value to the teacher in the rural and ungraded school" (p. ii). In many such schools, the teacher was responsible for several grades simultaneously, so the time that the teacher could spend with each grade was extremely limited. Although Alberta had embarked on a program of eliminating one-room rural schools through centralization during World War II, the process took many years to complete. LaZerte's observations continued to be valid for several years (Alberta Department of Education, *Annual Reports*, 1939-1965). The use of the *Numbers Tell Their Story* program appears to be ideally suited for such environments. LaZerte (1953) noted that "if he [the student] is to profit from the course, each pupil must do all the work himself. The teacher must not teach demonstration lessons. It is assumed that all pupils will have some equipment in their hands while the class is at work" (p. ii).

It should be recalled that LaZerte's desire for such an individualized program may be observed as early as 1922 (LaZerte, 1922, p. 30). At that time, there were some prominent psychologists who believed that practical and concrete methods should be used in the instruction of concepts. Thorndike (1923) proposed three criteria for a successful program:

> (1) Provide enough actual experiences before asking the pupil to understand and use an abstract or general idea. (2) Develop such ideas gradually, not attempting to give complete and perfect ideas all at once. (3) Develop such ideas so far as possible from experiences which will be valuable to the pupil in and of themselves. (p. 179)

LaZerte's (1953) *Numbers Tell Their Story* program appears to have met all three of Thorndike's criteria. To be sure, much of the material mentioned above, such as bead frames and felt cut-outs, were already available from school suppliers. Indeed, some of the items are still available (Primary School Supplies, 1987). Some of the apparatus was designed by LaZerte, however, and was not already available. Such equipment included perception cards (domino-like flash cards); number board (a wooden tray which could hold colored blocks in order to actualize addition and subtraction); place value frame (a shallow cardboard or wooden box marked in thirds, with nine slots in each third designed to hold colored disks); and the hundred board (a 23-inch square board, marked in quadrants with 25 hooks or nails arranged in rows in each quadrant, to illustrate submultiples of 100) (LaZerte, 1953; LaZerte, Dey, & Svidal, 1959). Originally, if one desired, complete sets of equipment could be ordered from the publisher (LaZerte, 1953, p. iv). In addition, concise instructions were provided on how to fabricate all the equipment mentioned in the book. As LaZerte (1953) noted, "if a school feels that it cannot undertake the expense of one set per child, the laboratory equipment is of such a simple nature that much of it can be reproduced in the school itself, at little cost" (p. iv). LaZerte's economical laboratory program was unlikely to be rejected because of its costly equipment, a consideration that did not seem to occur to Skinner and others who promoted expensive teaching machines and equally expensive instructional courseware.

Although prepared laboratory equipment packages were not offered from the publisher in later editions of *Numbers Tell Their Story*, the book remained in print until the mid-1960s (royalty statements in the LaZerte Papers, University of Alberta Archives). From this evidence, it appears that LaZerte's innovation of laboratory mathematics was reasonably successful, although it was eventually eclipsed by devices and curricula developed by individuals who took advantage of the mass-market distribution of publishers and manufacturers. By such means, many inappropriate and inferior devices and programs were, and still are, being foisted on educators as well as the general public.

Implications

In 1959, several years after his retirement, LaZerte wrote an article in which he foresaw little positive change in education and the repetition of previous mistakes because of a lack of innovation by educators. LaZerte (1959) believed that "too many of our colleagues are timid—afraid of change—afraid to use their professional freedom" (p. 49). LaZerte's observation is a likely explanation why various specialists in other disciplines have sought to improve education; it appears to them that little is being done within the discipline to improve it (Skinner, 1984). In addition, when it appears that the accomplishments of previous educators are being ignored or forgotten, others may fear that no progress has taken place.

The development and marketing of devices to enhance learning has continued during the last 50 years, with much impetus coming from the United States. While some of these devices have been useful in Canadian schools, others have been inappropriate because of their cost or incompatibility with the curriculum or school system (Rutherford, 1961). A review of the work of M.E. LaZerte indicates that as early as 1922, devices and materials to support learning were being developed and researched in Canada. Although some of LaZerte's research indicated that certain devices did not enhance learning in the most efficient manner, some of his devices and methods did, such as his *Numbers Tell Their Story* program, which was designed with Canadian schools in mind, particularly rural schools. LaZerte's work indicates that Canadian educators are sufficiently competent and innovative to engage in such research and development; they do not necessarily have to follow a lead or trend originating in the United States. Furthermore, LaZerte's experiments and findings suggest that there is a need for educators to evaluate critically the effectiveness of such devices and materials in the context of the local classroom before they are widely accepted for use in the field of education.

Acknowledgments

The author gratefully acknowledges the assistance provided by Dr. H.T. Coutts, former Dean of the Faculty of Education, University of Alberta, and that of Dr. S.M. Hunka of the Division of Educational Research Services, University of Alberta.

References

Alberta Department of Education. (1939-1965). *Annual reports*. Edmonton: King's/Queen's Printer.

Chalmers, J.W. (1978). *Gladly would he teach: A biography of Milton Ezra LaZerte*. Edmonton: The ATA Educational Trust.
Charyk, J.C. (1974). *Pulse of the community: Volume II of the little white schoolhouse*. Saskatoon: Western Producer Book Service.
Hendershot, C.H. (1964). *Programmed learning: A bibliography of programs and presentation devices*. Bay City, MI: Author.
Johns, W.H. (1981). *A history of the University of Alberta: 1908-1969*. Edmonton: University of Alberta Press.
LaZerte, M.E. (1922). Elementary mathematics. *The ATA Magazine Easter Annual*, 30.
LaZerte, M.E. (1926). *A study of the methods used by elementary school pupils in solving problems in arithmetic*. Edmonton: Unpublished Bachelor of Education thesis.
LaZerte, M.E. (1927). *A diagnosis of diffulties encountered in solving problems in arithmetic*. Chicago: Unpublished doctoral dissertation.
LaZerte, M.E. (1933). *The development of problem solving ability in arithmetic: A summary of investigations*. Toronto: Clarke, Irwin.
LaZerte, M.E. (1945). *I love freedom*. Edmonton: Unpublished speech notes.
LaZerte, M.E. (1951). *Arithmetic in the primary grades*. Edmonton: Unpublished speech notes.
LaZerte, M.E. (1953). *Numbers tell their story*. Toronto: Clarke, Irwin.
LaZerte, M.E. (1959, May 6-10). The road ahead. *The ATA Magazine*, 49.
LaZerte, M.E., Dey, J.D., & Svidal, R. (1959). *Numbers tell their story: Grade two teacher's manual*. Toronto: Clarke, Irwin.
Lumsdaine, A.A., & Glaser, R. (Eds.). (1960). *Teaching machines and programmed learning: A source book*. Washington, DC: National Education Association.
Pressey, S.L. (1932). A third and forth contribution toward the coming "industrial revolution" in education. *School and Society, 36*, 668-672.
Primary School Supplies. (1987). *School catalogue*. Richmond Hill, ON: Author.
Quintilian. *Institutio oratoria, I* [Training in oratory, I] (H.E. Butler, Trans., 1921). London: William Heinemann.
Rutherford, G. (Ed.). (1961). *Programmed learning and its future in Canada*. Ottawa: Canadian Teachers' Federation.
Skinner, B.F. (1954). The science of learning and the art of teaching. *Harvard Educational Review, 24*, 86-97.
Skinner, B.F. (1983). *A matter of consequences: Part three of an autobiography*. New York: Knopf.
Skinner B.F. (1984). The shame of American education. *American Psychologist, 39*, 947-954.
Thorndike, E.L. (1923). *The psychology of arithmetic*. New York: Macmillan.
Vargas, J.S. (1986). Instructional design flaws in computer-assisted instruction. *Phi Delta Kappan, 67*, 738-744.

The Alberta Journal of Educational Research *Vol. XXXV, No. 2, June 1989, 123-133*

JOHN O. ANDERSON
University of Victoria

Evaluation of Student Achievement: Teacher Practices and Educational Measurement

The evaluation of student achievement is a characteristic element of formal classroom education. Educational measurement addresses issues of and develops procedures for student evaluation. Recently there has been an increase in expressed interest in the relationship between educational measurement and the classroom practices used in evaluating student achievement (Stiggins, 1988). Questions have been raised as to the utility of procedures developed through educational measurement to teaching and learning. There appears to be a disparity between the science of measurement and the craft of evaluation as practiced in classrooms.

This study attempted to shed some light on classroom evaluation practices by considering whether testing is a common classroom practice, what kinds of student abilities or characteristics should be measured, what methods are used to collect this information, and what is the extent to which external sources of instrumentation or data are of interest to teachers. The 1986 British Columbia Provincial Science Assessment (Bateson, Anderson, Dale, McConnell, & Rutherford, 1986) provided data that could address these issues. This study reports the data derived from the teacher questionnaires in regard to the student evaluation practices and attitudes of science teachers in British Columbia and discusses some implications for the improvement of applied educational measurement.

The evaluation of student achievement is a characteristic element of formal classroom education. Tests, assignments, and projects are developed, administered, and marked by teachers in order to generate information about the achievement of individual students. Educational measurement is a field of development and research which addresses issues of and develops procedures for the evaluation of student achievement. These issues include test reliability, validity, standard error of measurement, item bias, test equating, and models underlying response data. A large number of formalized procedures have been developed for test design and construction, item analysis, test scoring and analysis, and standard setting.

The relationship between educational measurement and classroom practice should be one of utility: Educational measurement should produce information and procedures which are useful in the evaluation of student achievement. Recently there has been an increase in expressed interest in the relationship between educational measurement and the classroom practices used in evaluating student achievement (Stiggins, 1988). Questions have been raised as to the utility of procedures developed through educa-

---

John Anderson is an assistant professor of Educational Psychology at the University of Victoria. He is a specialist in student assessment.

tional measurement to teaching and learning. There appears to be a disparity between the science of measurement and the craft of evaluation as practiced in classrooms.

Stiggins, Conklin, and Bridgeford (1986) noted in their extensive review of the literature on classroom assessment that the dominant paradigm in educational measurement is based on data generated from standardized paper-pencil tests used for the purposes of accountability. This focus of educational measurement is well illustrated in a recent description of the future of testing (Rudman, 1987). The article concentrates solely on published, standardized tests and does not mention the development and use of classroom tests by teachers. In contrast, Gullickson (1986) notes in his American study of teacher-perceived needs in measurement that teachers would favor more emphasis on the use of nontest methodologies in measurement and evaluation courses. These methodologies include rating scales, observation, and anecdotal records. Teacher interest in measurement is directed toward a variety of instruments and procedures developed for instructional purposes and not toward the issues and procedures of mainstream educational measurement research and literature.

As a specific example of the disparity between measurement and instruction, consider a recent critique of the reporting of results of the National Assessment for Educational Progress (NAEP) project in the United States (McLean & Goldstein, 1988). It was noted that, after a great deal of complex analysis involving large data sets collected over several years, NAEP was able to report reading results on a single scale. McLean and Goldstein point out that this modeling of reading ability as a single trait does not coincide with conceptions of reading abilities that are current among those involved in the teaching of reading. The measurement project (NAEP) produces results that are meaningful within the measurement perspective of unidimensional human traits, but this perspective does not fit the multidimensional view of those involved in the study and teaching of reading.

Although the focus of educational measurement is on large-scale testing projects and standardized tests, it appears that most of the achievement testing done is in classrooms by teachers as part of ongoing evaluation of student achievement. An estimate of the number of tests a student completes during an academic career was provided by the results of a questionnaire administered by the author in the fall of 1987 to a group of 92 fourth- and fifth-year University of Victoria students of an educational measurement course. The students were asked, "How many formal tests or examinations (those counting for grades) would you estimate you have taken in your life?" The median response range was 500 to 600 such tests. The vast majority of these tests would not be standardized tests but rather tests developed by classroom teachers. In British Columbia, for example, a province that does have a provincial grade 12 examination program, a grade 12 scholarship exam program, and a provincial learning assessment program, most students continuing through to postsecondary levels would write fewer than 20 such tests in their public school (K to 12) career. It is likely, then,

that most tests a student encounters are developed by classroom teachers to assess student achievement in specific instructional programs.

The purpose of this study was to explore a facet of the relationship between educational measurement and the evaluation of student achievement. Because it appears that most of the testing to which students are exposed over their educational careers consists of procedures and instruments developed by classroom teachers to assess student performance in instructional programs, a first step in exploring the utility of educational measurement is to investigate what goes on in the classroom. This could lead to the determination and perhaps the improvement of the fit between educational measurement and classroom practice.

This study attempted to shed some light on classroom evaluation practices by considering whether testing is a common classroom practice, what kinds of student abilities or characteristics should be measured, what methods are used to collect this information, and what is the extent to which external sources of instrumentation or data are of interest to teachers. The 1986 British Columbia Provincial Science Assessment (Bateson, Anderson, Dale, McConnell, & Rutherford, 1986) provided data that could address these issues. The 1986 British Columbia Science Assessment investigated the current state of science education in the province through the administration of achievement measures and attitude scales to all students in grades 4, 7, and 10, and the administration of questionnaires to all teachers of grades 4, 7, and 10 science. Data collection took place in the spring of 1986.

The teacher questionnaires focused on classroom practices and teacher attitudes toward various aspects of their teaching: teacher background and general information, coordination of science education, physical facilities, materials and equipment, science teaching, assessment and evaluation, and the science program. The grade 10 questionnaire contained 56 items. Grades 4 and 7 teachers were administered identical 34-item questionnaires.

This study reports the data derived from the teacher questionnaires in regard to the student evaluation practices and attitudes of science teachers in British Columbia. In total 1,895 grade 4 teachers, 1,519 grade 7 teachers, and 1,035 grade 10 teachers responded to the questionnaire. The grades 4 and 7 teachers were essentially classroom teachers, teaching science to their homeroom class in addition to teaching a broad range of other subjects to these students. Grade 10 teachers tended to be specifically assigned to teaching science (Bateson et al., 1986).

These data provide a start for a study of teachers' practices in and attitudes toward student evaluation. It is in this light that this article has been prepared. It is not to be inferred that these data fully answer the questions posed. However, the responses to the limited number of questionnaire items by science teachers in British Columbia suggest directions the answers would take if a broader sample of teachers were asked a broader sample of questions and if a more comprehensive set of methodologies were employed.

Results

In responding to the question: "Is testing a common classroom practice?" teachers of grades 4 and 7 science were asked to indicate how often 10 classroom activities were carried out in a typical week. The results (Table 1) indicate that, of those activities listed, testing was the most frequent classroom activity, with approximately 70% of teachers reporting that students were tested "Always" or "Often" each week. Testing was a more frequent classroom activity than conducting science demonstrations or student experiments, using the computer, or doing homework. The pattern of responses was similar for both grades 4 and 7 teachers.

To answer the question of whether testing is a characteristic activity of the classroom, it is clear that testing is a common classroom activity.

In considering what kinds of student abilities or characteristics should be measured, grade 10 teachers were asked about the extent to which each of the four science goal areas—Attitude, Skills and Processes, Knowledge, and Critical Thinking (Ministry of Education, 1981)—should be tested in the determination of student achievement. In general, teachers indicated that all four areas should be tapped for student evaluation purposes.

Attitudes were considered to require less emphasis than the other areas, with over one fifth of the teachers indicating that Attitudes should not be

Table 1
Science Program – Activities in the Classroom
Grades 4 and 7

Item	Percentage of Teachers							
	Grade 4				Grade 7			
In a typical week of science classes, how often did you do the following?	A	O	S	N	A	O	S	N[1]
Carried out experiment for students	1	10	77	12	0	11	73	16
Handed out notes	1	12	52	35	0	13	56	31
Wrote out notes for students	2	22	60	16	2	23	62	13
Students conducted experiments individually	2	16	70	12	2	22	66	10
Students conducted experiments in small groups	2	19	64	15	2	31	55	12
Students used computer	0	2	15	83	0	3	19	77
Students used library books	5	25	67	3	2	14	76	9
Students chose topics to study	2	3	52	44	0	6	46	48
Students did homework	2	13	74	12	6	35	56	3
Students were tested	36	33	30	1	39	36	24	0

[1] A = Always, O = Often, S = Sometimes, N = Never

Table 2
Assessment and Testing – Curricular Goals Emphasis
Grade 10

Item	Percentage of Teachers
In the determination of the final grade, what emphasis should be placed on the following science goal areas?	
Attitudes	
0%	22
10%	56
20%	19
30%	2
40%	1
50%	0
60%	0
Skills and Processes	
0%	1
10%	10
20%	31
30%	44
40%	12
50%	3
60%	0
Knowledge	
0%	0
10%	1
20%	8
30%	23
40%	28
50%	22
60%	19
Critical Thinking	
0%	1
10%	22
20%	46
30%	23
40%	6
50%	1
60%	1

used at all in evaluating students. A majority of teachers (56%) indicated that Attitudes should count for 10% of the final grade. Essentially all teachers indicated that Skills and Processes, Knowledge, and Critical Thinking should receive emphasis in the determination of a science student's final grade. The response modes were Skills and Processes, 30% emphasis; Knowledge, 40% emphasis; and Critical Thinking, 20% emphasis. In summary, science teachers regard a number of different student skills and characteristics to be of importance for evaluation purposes. The weighting applied to the skills and characteristics varies from one category to another with Knowledge receiving the heaviest weighting and Attitudes the least.

Science teachers were asked about the methods they used for evaluating student achievement. A number of common instruments and procedures were listed and the teachers were asked to indicate the emphasis they placed on each for student evaluation purposes. Table 3 lists the response frequencies for grades 4 and 7 teachers in relation to types of methods used to collect information about student achievement. Table 4 lists the response frequencies for grade 10 teachers.

Table 3
Assessment and Testing—Source of Assessment Data
Grades 4 and 7

Item	Percentage of Teachers							
	Grade 4				Grade 7			
What emphasis do you place on each of the below in deriving a final grade for a student in your science class?	N	L	S	M	N	L	S	M[1]
Anecdotal records	7	24	58	11	16	32	46	6
Teacher-made objective tests	1	3	48	48	0	2	43	54
Ministry of Education Classroom Achievement Tests	62	22	15	1	58	22	18	1
Subjective tests	14	22	58	7	9	23	58	10
Activity/experiment write-up	5	16	62	16	4	11	58	27
Individual work contracts	53	13	28	7	54	15	24	8
Reports on topics in science	6	15	68	11	6	17	64	14
Projects	4	9	70	17	4	14	64	19
Oral tests	35	28	33	5	47	31	20	2
Student self-reports	52	21	22	5	54	21	23	2
Attendance	67	13	14	6	67	17	9	6

[1] N = No emphasis, L = Little emphasis, S = Some emphasis, M = Much emphasis

Table 4
Assessment and Testing – Source of Assessment Data
Grade 10

Item	Percentage of Teachers		
How much importance do you place on each of the following in deriving final student evaluation?	Very Important	Important	Not Important
Anecdotal records			
– achievement	19	39	41
– general attitude	4	48	48
– work habits	8	63	29
Teacher-made objective tests	71	28	1
Standardized objective tests	14	43	43
Subjective tests	30	56	14
Laboratory write-ups	37	59	3
Individual work contracts	6	21	72
Reports on topics	17	64	20
Projects	17	61	22
Oral tests	4	26	70
Student self-reports	3	23	75
Attendance	17	50	33
Classroom behavior	15	43	42

Two trends emerge. First, teachers used a wide variety of methods to collect information related to student achievement: anecdotal records, subjective tests, experiment write-ups, reports, and projects. Second, in all grades, teacher-made objective tests received the most emphasis (and presumably use) in the determination of student achievement. The Ministry of Education Classroom Achievement Tests, individual work contracts, oral tests, and student self-reports appear to be the least emphasized forms of assessing student achievement. The majority of grades 4 and 7 teachers (67%) did not use attendance for the determination of final grades. Yet most grade 10 teachers (67%) consider attendance to have at least some importance in the determination of a student's final grade.

In summary, teachers do not rely on a single source of data for evaluating student achievement. They use a variety of methods to measure student performance with teacher-made objective tests receiving the most emphasis of those methods presented.

Grade 10 teachers were asked about their preferences for using school, district, or provincial examinations for evaluation of student achievement.

The questions were asked in order to gauge teacher interest in and possible use of external sources of instrumentation. Teachers were asked the extent to which these various tests should "count" toward a student's final grade (Table 5). There was considerable support for the use of a schoolwide exam. A majority of teachers (79%) thought that student assessment should include the results of a schoolwide test counting for 20% or 30% of the final mark. However, 21% of teachers indicated that a schoolwide exam should not be used at all. In regard to the use of district or provincial exams, approximately 75% of grade 10 science teachers indicated that they would not like to base student assessment on districtwide or provincial exams. It ap-

Table 5
Assessment and Testing—Instrumentation for Student Grade
Grade 10

Item	Percentage of Teachers
Would you like to have student assessment based on ...	
schoolwide exam?	
No	21
Yes, about 10%	8
Yes, about 20%	50
Yes, about 30%	13
Yes, about 40%	4
Yes, about 50%	4
districtwide exam?	
No	75
Yes, about 10%	10
Yes, about 20%	12
Yes, about 30%	2
Yes, about 40%	0
Yes, about 50%	2
provincial exam?	
No	77
Yes, about 10%	6
Yes, about 20%	9
Yes, about 30%	5
Yes, about 40%	1
Yes, about 50%	3

pears that teachers did not favor the use of either district or provincial level tests for student evaluation purposes. It should be noted that a teacher would likely have direct involvement with the development of a school exam, whereas the opportunity for direct involvement is somewhat less for a district exam and almost nil for a provincial exam.

Teachers of grades 4 and 7 science were asked whether they had read any of the reports of the previous Provincial Science Assessment (Taylor, 1982) in order to estimate their interest in external sources of data related to student achievement in science (Table 6). Reports had been produced at the provincial and district levels; no school or class level reports were produced in this provincial assessment. Most teachers (approximately 70% or more) indicated that they had not read any of the reports produced. These results suggest that external sources of data, in this case from a provincewide assessment of student performance in the subject they are teaching, are not of great interest or importance to teachers.

In regard to the question of teacher use of external testing, the data presented here indicate that science teachers do not view positively the use of external testing for student evaluation. Nor do they have much interest in the data generated from large-scale assessments.

Discussion

The results suggest that testing is a common classroom activity and that there is substantial variation in the methods and formats used to estimate student achievement. Teacher-made objective tests are the most common form of instrumentation in use, although many other formats for data generation and collection are used by science teachers in British Columbia. Teachers believe that student achievement should be evaluated in a number of different domains: attitudes, knowledge, and skills. Further, teachers are not interested in the use of external forms of testing, either as instrumenta-

Table 6
Assessment and Testing—
Utilization of 1982 Science Assessment Results
Grades 4 and 7

Item	Percentage of Teachers Responding *Yes*	
	Grade 4	Grade 7
Have you read the following publications of the 1982 Provincial Science Assessment?		
District Interpretation Report	25	32
Provincial Update flier	15	21
Provincial Summary Report	25	32
Provincial General Report	17	22

tion or as the results of large-scale assessment, even in a subject they are currently teaching.

The results pertaining to teacher test use are not surprising. In an earlier study of a broad sample of Ontario teachers (Wahlstrom & Danley, 1976), it was found that teachers used a variety of methods in measuring student achievement and preferred to use instruments they developed themselves. However, Wahlstrom and Danley found that the preferred assessment technique of Ontario elementary teachers was observation of classroom work by the teacher, whereas secondary teachers prefer tests. These findings were similar to a recent American study (Stiggins & Bridgeford, 1985) which also found variation in preferences for student evaluation procedures by teachers across subject areas as well as across grade levels.

The pervasiveness of testing in the classroom coupled with the use of a wide variety of formats to measure different kinds of student achievement does raise a number of important issues in regard to the relationship between educational measurement and teacher practices in classroom assessment. To illustrate, consider one such issue: the combining of scores from different sources to arrive at a final grade. How do teachers combine the scores from a number of different testing formats when testing a number of different student abilities and characteristics collected over a period of time? Certainly the scores are combined, but how and with what effect? It is likely the task is conducted by simply summing up the numbers and assigning grades to these numbers. It would be of interest and utility to investigate the procedures actually employed in order to determine their consequences and the consistency of implementation within and between classrooms. Nagy, Traub, and MacRury (1986) have described methods for research into the procedures used by teachers in determining lettergrades (standard setting). A description of these procedures and patterns of implementation would also illuminate some of the practices used by teachers, not only in the combining of scores to convert to grades, but also the attitudes of teachers toward achievement, and an estimate of the variability of achievement standards. A number of such studies are in progress (MacRury, 1988; Nagy, 1988; Traub, 1988; Wilson & Rees, in press).

However, it is of concern that there is a widespread practice in student assessment that is not well addressed by educational measurement. It appears that there is a substantial practice of combining test data from various sources, but the issue is largely ignored in the literature. On the basis of an informal review of educational measurement textbooks and journals by the writer and also as noted by Stiggins et al. (1986), the literature focuses on single-test situations, that is, the development, administration, and analysis of a single test (or set of tests) as a discrete project. The most common manifestations of the single-test situation in schools are final examinations, assessment programs (such as the one from which the data of this study were derived), and standardized test programs—areas that are of importance educationally and lend themselves to systematic research and development, but are of little interest or use to the classroom teacher (Table

6). Yet it is likely that for most teachers the measurement practices employed in the classroom are derived from personal experience both as teachers and as students, and from techniques garnered from colleagues. These classroom practices would form the basis of "the craft of evaluation" (McLean, 1985) and would not necessarily coincide with practices derived from the science of measurement and its literature.

Assuming that most of the achievement testing being done in Canada is that done by teachers in classrooms, it appears that this practice could be better served by research. The apparent isolation of classroom testing practice from research activity is a problem that should be addressed to improve overall educational practice. The determination and description of the measurement practices in current use and the extent of variation in practice could lead to the development of better techniques for handling the data generated by classroom measurement practices and the generation of better data by classroom measurement. Because classroom derived technique predominates testing and student evaluation practices, applied educational measurement should center its attention on the classroom.

References

Bateson, D.J., Anderson, J.O., Dale, T., McConnell, V., & Rutherford, C. (1986). *The 1986 British Columbia science assessment: General report.* Victoria: British Columbia Ministry of Education.

Gullickson, A.R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement, 23,* 347-354.

MacRury, K. (1988). *Assigning course grades: Policies and practices.* Paper presented to the annual meeting of the National Council on Measurement in Education, New Orleans.

McLean, L.D. (1985). *The craft of student evaluation in Canada.* Toronto: Canadian Education Association.

McLean, L.D., & Goldstein, H. (1988). The U.S. national assessments in reading: Reading too much into the findings. *Phi Delta Kappan, 65,* 369-372.

Ministry of Education. (1981). *Elementary science curriculum guide: Grades 1-7.* Victoria: British Columbia Ministry of Education.

Nagy, P. (1988). *The content basis of teaching and testing in mathematics.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.

Nagy, P., Traub, R.E., & MacRury, K. (1986). *Strategies for evaluating the impact of province wide testing.* Toronto: Ontario Ministry of Education.

Rudman, H.C. (1987). The future of testing is now. *Educational Measurement: Issues and Practices, 6*(3), 5-11.

Stiggins, R.J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 65,* 363-368.

Stiggins, R.J., & Bridgeford, N.J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22,* 271-286.

Stiggins, R.J., Conklin, N.F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practices 5*(2), 5-17.

Taylor, H. (1982). *The 1982 British Columbia science assessment: General report.* Victoria: British Columbia Ministry of Education.

Traub, R.E. (1988). *A study of the marking process: Calculus teacher responses to a common marking task.* Paper presented to the annual meeting of the National Council on Measurement in Education, New Orleans.

Wahlstrom, M.W., & Danley, R.R. (1976). *Assessment of student achievement.* Toronto: Ontario Ministry of Education.

Wilson, R.J., & Rees, R. (in press). The ecology of evaluation: Measurement and evaluation within an administrative framework. *Canadian Journal of Education.*

The Alberta Journal of Educational Research *Vol. XXXV, Number 2, June 1989, 134-144*

ROBERT J. WILSON
Queen's University

Evaluating Student Achievement in an Ontario High School

Although classical test theory has been a productive force in educational research for almost a century, its impact on classroom practice in evaluation has not been especially strong. Little is known about what teachers actually do (as opposed to what they report doing). The policy and procedural context in which the evaluation of students occurs is similarly unexplored. The present study was conducted to examine what teachers do in the context of their classes and their school's policies. Thirty-six teachers in a high school collected all their evaluation instruments for one class for a semester while maintaining a log and responding to a questionnaire for each instrument. Descriptive statistics for such categories as frequency of evaluation activity and purpose were computed. The results are discussed in terms of the complexity of the tasks these teachers faced in their evaluation practice.

One of the few areas of research in education that has developed in a classical, theory-driven manner is the evaluation of educational performance. Test theory, established over three quarters of a century, is presently alive on several fronts with empirical research. Examples include work in item response or latent trait (Hambleton & Swaminathan, 1985; Lord, 1980) and generalizability theory (Brennan, 1983; Cronbach, Gleser, Nanda, & Rajaratnam, 1972).

Despite this activity, the relationship of research-based advances to school practice in student assessment has been and remains tenuous. Most evaluation texts and materials prepared for practicing educators adapt the test-building principles used by professionals (e.g., domain specification, test blueprinting, item format selection, piloting, item analysis, reliability, and validity checks) and assume that these adaptations, if made, would help teachers construct useful measures for a variety of classroom purposes (Brown, 1981; Cunningham, 1986; Thorndike, 1982). Yet one investigator (McLean, 1985), reviewing the "craft" of student evaluation in Canada, concluded:

> The elemental stuff of the measurement courses and textbooks is as foreign to classroom teachers as spectral analysis of glaze mixture is to potters—and apparently as irrelevant. (p. 39)

One reason why this gap may exist is that test theory concerns itself primarily with issues of measurement (hence psychometrics). Because the purposes to which any instrument may be put are largely in the control of the users, the professional test producer can only anticipate the most com-

---

Robert Wilson is a professor of education at Queen's University. His area of interest is measurement and evaluation.

mon ones and provide as much relevant validity evidence as possible for those uses. As for the utility of the measure in promoting more effective educational decision making, the professional test producer is even more removed. Thus the conceptual model for professional test producers may be summarized as a fundamental concern with the technical characteristics of the measures (particularly as these affect the reliability of the scores); a large concern with validity in terms of anticipated usage (even though that usage is beyond the producer's control); and a less-than-dominant concern for the overall utility of the measure in the various contexts in which it may be used.

While all three of these aspects are likely to be present in school practice as well, their configuration will be quite different. Teacher-constructed instruments are more likely to begin with decisions. Before teachers invest the time and resources necessary to develop a test, they will probably be cognizant of at least one specific need that the test is to serve. Because teachers are also more likely than test producers to know their clients, concrete expectations about performance will affect the construction of the instrument itself. (See Wilson & Rees, in press, for an elaboration of this model in more theoretical terms.)

Consequently, a sensible order of events for a teacher-prepared test is likely to be the reverse of that of the professional, and the test preparation process will differ as well (Gullickson, 1986). The concern of the teacher will be centered on the decisions to be made from the instrument (the production of grades for reporting purposes and the documentation of work completed are common examples) and relatively less on the characteristics of the measure itself. Much of this analysis is speculative, however, as little study beyond surveys of real behavior or the concepts which shape it has been conducted with teachers (Stiggins, Conklin, & Bridgeford, 1986).

The research evidence that is available tends to partition itself into two categories. The first assumes that research findings could be most useful if they delineated what criteria and standards should be evident in classroom practice. In the inaugural issue of the *Journal of Educational Measurement*, Mayo (1964) documented the need for teacher understanding of such fundamental measurement concepts as reliability and validity. How the needs were documented reflected a particular approach to improving teacher practice. Mayo identified instructors of measurement courses (as well as some others) and had them rate various statements concerning knowledge in the area in terms of perceived relevance to teachers. On this basis, presumably, courses and texts could be developed that would meet these needs.

More recently, Rudman, Kelley, Wanous, Mehrens, Clark, and Porter (1980) continued to find large gaps in American teachers' understanding of these concepts. Frisbie and Friedman (1987), accepting the gaps as given and as undesirable, have attempted to define the fundamental needs for classroom teachers in terms of an analysis of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985). Through the use of the standards, they argue, a domain description outlining the knowledge

structure fundamental to sound evaluation practice could be developed. Once developed, this structure could form the basis of teacher education in this important area.

A second category of research, more recent in origin, attacks the issue from the perspective of the teacher. Researchers such as Gullickson (1984) argue that technical adequacy, for example, is only one of several factors influencing teachers' use of tests. Student acceptance and support of the testing model, the nature and quality of the information to be obtained from the test, external constraints, and the way in which the teacher intends to use test results may also be important factors influencing evaluation practice from the teacher's point of view.

In a series of studies by Stiggins and his colleagues (Stiggins & Bridgeford, 1985; Stiggins, Conklin, & Bridgeford, 1986), the issue of the applicability of the typical courses provided teachers in educational measurement has been raised. Teachers surveyed in these studies, for example, seem to rely most on the very measures (observation, assignments) that evaluation specialists relegate to minor status in their courses. The same teachers tend to view the results of standardized tests as only minimally useful, yet courses in measurement tend to stress the administration and interpretation of these measures far more than the preparation and interpretation of the techniques teachers use most often (Stiggins, Conklin, & Bridgeford, 1986).

Although the issue of context is critical in the pursuit of this second approach to research, few studies from this second perspective have collected the instruments teachers use. Thus far, most of the effort has been directed at results obtained from questionnaires. Secondly, there may be a need now to place the thoughts and actions of teachers into a more formal pattern. Perhaps teachers do have theories which guide their practice even if those theories are not fully developed, expressed, or matched to those of the evaluation professionals.

The following investigation was pursued (a) in an effort to examine what some educators think and do in the context of their classes and (b) from this examination, to develop a model that might illuminate those actions.

Method

The investigation occurred in a high school situated in a relatively homogeneous, middle-class, suburban area of eastern Ontario. The programs offered to approximately 600 students include "general" as well as "advanced" subjects, but the emphasis in the school is on academic courses. Typically, over 80% of the credits awarded will be of an advanced level. In any given year, 70% of this school's graduates will enroll in postsecondary institutions.

The evaluation study conducted in this high school resulted from a concern by a new principal that grades reported to parents be comparable from one subject to another. For example, the mathematics department, because of its commitment to a form of criterion-referencing, was typically producing negatively-skewed distributions while others, particularly those in the humanities, were generating more symmetrical ones. A committee com-

posed of the principal, the author, and two teachers was charged with the responsibility of investigating what policies and procedures of evaluation would be most appropriate for the school. As part of this process, a survey of present practice was conducted. It is this survey that is reported on here.

The Survey

As part of the process of developing policies and procedures for the school, the committee decided to find out what evaluation practices were presently in force. Each of the 36 full-time teachers in the school collected every evaluation tool he or she used for one course for a full semester.

All teachers participated in the data collection phase of the survey. Each was assigned one class for which the following data were required: an evaluation log documenting each instrument's title, its date of use, the number of students who completed it, the numbers of As and Fs awarded, the median, total marks, and the number of classes participating; a copy of the instrument; and an evaluation checklist indicating the purposes of the instrument (from seven choices offered); an assessment of the achievement of the class in terms of teacher expectations; and any explanatory comments by the teacher concerning the context.

Results

Frequency of Evaluation

The frequency with which students were evaluated was a primary concern of the school in terms of efficient use of instructional time and adequacy of assessment. (Table 1 indicates the total number of assessments per class as well as the frequency certain purposes were cited.)

These data show that there was a large degree of variability from course to course in terms of the number of instruments used. However, for a typical student averaging 107 hours of instruction per course in this semestered school, some evaluation would occur approximately every nine hours. For such a student carrying a full program of three to four courses, some type of formal evaluation activity was occurring approximately every second day of school.

The frequency with which evaluation occurred could be related to other factors such as the subject matter and the level at which the courses were being taught. Table 2 displays the information concerning level and subject, with subject being dichotomized into "science and science related" and "humanities." These two perspectives do not seem to account for the variability observed in the data, thus indicating that it is the individual teacher whose view dictates the frequency of evaluation no matter what the course or level.

This conclusion is reinforced by examining the correlations between frequency and purposes of evaluation displayed in Table 3. Those teachers who checked most often that their purposes for evaluating students included checking students' progress against course objectives (Crit), generating marks for reporting purposes (Mark), diagnosing students' weaknesses with material (Diag), and having students practice or apply what has been learned (App) were also those who tended to have more frequent evalua-

Table 1
Frequency of Evaluation by Subject and Purpose During One Semester

Teacher	Purposes							
	Crit	Mark	Diag	App	Cons	Norm	Next	Total
Science and Science Related								
Auto Mechanics	1	3	0	3	2	1	0	3
Biology	53	53	52	46	8	11	0	53
Business	3	3	0	2	3	2	0	3
Business	11	10	8	9	7	7	0	12
Business	2	7	1	12	3	1	1	14
Business	3	0	0	3	2	0	0	3
Chemistry	4	4	2	4	4	1	0	4
Computers	10	10	7	8	5	0	0	10
General Science	7	11	6	5	9	0	0	11
Mathematics	7	1	3	8	5	2	0	9
Mathematics	9	15	13	12	9	6	1	15
Mathematics	7	8	7	8	8	1	0	9
Mathematics	13	0	6	3	12	1	0	13
Mathematics	12	14	12	0	0	0	12	14
Physics	13	14	6	18	17	6	6	22
Physics	1	0	1	1	1	0	0	1
Woodworking	2	2	1	2	1	1	0	2

tions. Indeed, the high intercorrelations among these four categories indicate that they could be more adequately thought of as elements of a single general approach to evaluation.

Objectives-Based Versus Norm-Referenced

The orientation reflected in the tendency to give many evaluations often produces a negatively-skewed distribution of marks. The school policy, however, required a more symmetrical view of what mark distributions should be. To ensure consistency of grading within the school, the following guidelines had been issued by the principal to influence the reporting of grades:

1. Course medians should normally be in the 66%-72% range at all grades and levels of difficulty.
2. Failure rates should normally not exceed 15% in each course at all grades and levels of difficulty.

Table 1 (continued)

Teacher	Purposes							
	Crit	Mark	Diag	App	Cons	Norm	Next	Total
Humanities and Social Sciences								
Art	8	0	2	9	8	0	0	9
Drama	8	8	7	12	11	11	2	12
English	22	10	11	13	15	1	0	25
English	14	14	15	18	12	0	2	26
English	18	21	16	21	22	4	11	23
English	9	13	11	16	17	7	6	19
English	12	19	6	14	14	19	3	19
English	5	11	5	8	7	0	0	13
Family Studies	12	3	3	12	8	2	0	15
French	5	11	6	9	9	0	0	12
French	15	9	14	15	15	4	1	18
Geography	7	8	5	3	7	3	3	8
History	8	9	6	4	8	3	0	10
History	9	9	10	11	10	3	0	11
History	7	6	7	6	7	2	0	7
Music	10	10	10	10	10	10	4	10
Phys. Ed.	1	1	1	2	1	1	0	2
Phys. Ed.	10	10	0	6	0	0	0	10
Phys. Ed.	1	1	0	1	1	0	0	1

3. The number of first-class marks should normally equal or exceed the failure rate at all grades and levels of difficulty.

This conflict in needs between administrative and instructional levels in the system was one of the main reasons for the total investigation, but of more importance to the present discussion was how the teachers were able to coordinate the contradictory purposes of "checking students' progress against course objectives," a goal more suitable to objectives-based assessment, with another goal of reporting grades where students' achievement would be compared to others, a norm-referenced approach.

One common solution was to use more frequent but less heavily weighted evaluations for the objectives-based purposes. These instruments typically had few failures (Fs) and many high marks (As). As the time for mark reporting drew near, more heavily weighted but less frequent tests and examinations were used to bring the total mark distribution into line.

Table 2
Average Usage per Teacher by Subject and Level During One Semester

	N	Overall	By Purpose					
			Crit	Mark	Diag	App	Cons	Norm	Next
Science and Science Related									
Academic	12	13.1	11.0	10.3	9.0	9.2	6.5	2.5	1.6
General	3	9.7	4.7	6.7	3.0	8.0	4.0	3.0	1.6
Mixed	2	6.0	6.0	6.0	4.0	5.0	3.0	0.5	0.0
Subtotal	17	11.6	9.3	9.1	7.4	8.5	5.6	2.4	1.2
Humanities									
Academic	12	12.7	9.3	9.8	6.8	8.8	9.2	4.1	1.4
General	4	18.8	13.3	11.8	11.0	15.5	13.0	2.3	3.3
Mixed	3	7.7	5.7	3.0	3.3	7.7	6.7	4.0	0.7
Subtotal	19	13.2	9.5	9.1	7.1	10.0	9.6	3.7	1.7
Overall Total	36	12.4	9.4	9.1	7.2	9.3	7.7	3.1	1.4

Note: Crit: criterion-referenced; Mark: mark generation; Diag: diagnostic; App: application; Cons: consolidation of knowledge; Norm: norm-referenced; Next: decide what to teach next

These instruments usually had more of a balance between As and Fs and medians within the range called for by the school's policy or lower. (No doubt there are instructional reasons for this pattern as well: As more work is completed in a term, the capacity for more involved, integrative, and thus potentially more difficult, assignments increases.)

The marks summary for one typical teacher is displayed in Table 4. Of the first six classroom-based assignments, totaling 80 marks, there were only five failures. The last six classroom-based assignments totaled 155 marks and produced 13 failures. The final examination had only three As as well as three Fs, and a median at the lower end of the acceptable range of marks. Given the situation of a student receiving the median score on each of these 13 assignments, that student's final mark would be 71%, an average grade within the 65%-72% guideline but one in which most of the assignments which had produced it had the negative-skewness characteristic of objectives-based measures.

Discussion

A most striking feature of the data concerning purpose is that evaluation instruments invariably served at least two purposes, and frequently more than two. Part of this multiplicity of purpose is necessitated by the environ-

Table 3
Means, Standard Deviations, and Correlations for Frequency and Purposes*

	Freq	Crit	Mark	Diag	App	Cons	Norm	Next
Freq	12.44 (9.65)							
Crit	**.93**	9.42 (8.96)						
Mark	**.89**	**.87**	9.11 (9.32)					
Diag	**.89**	**.93**	**.92**	7.22 (8.97)				
App	**.93**	**.87**	**.87**	**.87**	9.28 (8.33)			
Cons	**.57**	**.42**	**.53**	.38	**.56**	7.72 (5.41)		
Norm	**.45**	**.42**	**.53**	**.42**	**.53**	**.46**	3.03 (4.23)	
Next	.25	.15	.29	.20	.15	.38	**.46**	1.47 (2.95)

Note: Crit: criterion-referenced; Mark: mark generation; Diag: diagnostic; App: application; Cons: consolidation of knowledge; Norm: norm-referenced; Next: decide what to teach next. Correlations in bold are significant beyond the .01 level.
*Means and Standard Deviations on diagonal

ment set by school policy. Instruments under this system cannot be considered as discrete evaluations; they must fit into an emerging pattern resulting in conclusions of overall performance called grades, conclusions that involve normative judgment. Thus the professional testmaker's concern with clarity and singleness of purpose, a unitary index of reliability, and a view of validity based on the administration of a single instrument may not have much applicability in the complexity represented by this milieu.

While both professional testmaker and teacher may share a concern for reliability in measurement, validity in judgment, and utility of decision making, the teachers in this school at least were also manipulating additional pressures related to focus. The students, the institution, and the teacher as self-evaluator all brought different demands to the exercise of evaluation.

Focus on Students

The need to make decisions about student progress (both formatively and summatively) precipitates much of the evaluation activity within the school. This dominant concern is reflected in the purposes checked, many of which

Table 4
Summary of One Teacher's Marks Record

Assignment Number	Number of As	Number of Fs	Median (in %)	Total Marks
1	7	1	73	15
2	4	0	75	10
3	14	0	92	20
4	0	3	70	10
5	4	1	75	10
6	18	0	96	15
7	5	2	72	40
8	8	5	62	30
9	8	1	70	10
10	10	1	100	10
11	2	2	70	15
12	5	2	64	50
13	3	3	65	100

could be subsumed under the general category of monitoring achievement, of providing objectives-referenced assignments and norm-referenced tests that generate marks leading toward a final, summary assessment.

Focus on Institution

From the school's point of view, this summarizing decision making should occur in a consistent form across subject areas, grades, and levels of program. Thus the school shapes the structure of its reporting system through its policy statements concerning median ranges, failure rates, and symmetricality of score distributions.

The acceptance of this school policy influences the focus teachers are able to maintain with their students. Knowing they must provide individual assessments, day to day as well as end of semester, teachers create expectations for their students that, among other things, contribute to the grading structure set up for the school as a whole. As Table 4 illustrates for one teacher, assignments tended to be easier, less heavily weighted early in the semester leading to a greater reliance on summative, more heavily weighted evaluations as the end of the course approached. The mandating of a formal examination period at the end of the semester for all courses also encouraged this pattern.

Focus on Teacher

Finally, the data indicate that the teacher may also use the results to review the course itself. This attribute may explain why 58% of the instruments were checked by teachers as having as a purpose "to diagnose students'

weaknesses with material" (Diag) but only 12% of the instruments were seen as useful "to decide what to teach next (Next)."

The teachers may have mentally recorded items needed for review without using that information formally with the class or classes which generated the data for that particular conclusion. Such a review would likely shape the teacher's own standards of expected, subsequent performance both for himself or herself and the students. The instruments and their results, then, may have given some teachers feedback on their own performance, leading them to conclusions about what to do differently in the future.

Conclusion

While the data collected in this study are closer to the best evidence criterion than those typically collected through objective means, it is still unclear whether teachers view activities of evaluation in a coherent, holistic manner.

At present, further investigations are being conducted which elicit rational and contextual data from more than one site and from representatives of more than one level within the system. It is only in such a comparative situation, with data solicited both qualitatively and quantitatively, that firmer conclusions concerning the impact of levels, for example, could validly be seen to be a differential and powerful influence on the evaluation of student achievement.

Nevertheless, the present analysis does provide some putative responses to frequently posed but seldom answered questions such as why teachers seem reluctant to adopt ready-made instruments for classroom use and why they so seldom apply the models of evaluation represented in measurement texts. The ready-made instruments may perform one part of the teacher's purpose for evaluation but not another. Such tests might, for example, provide for some relevant comparative data but not fit into the pattern required by the teacher for grade reporting. Similarly, the professional's concern with reliability of scoring, item analysis, and systematic content coverage may also seem of only esoteric interest, and even an imposition, to teachers practicing the delicate balancing acts required by the different, sometimes contrary, forces operating on them already.

Acknowledgment

The author wishes to thank Marilyn Connock, Gary Medd, Ruth Rees, and two anonymous reviewers for their helpful comments on an earlier draft of this article.

References

American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME). (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Brennan, R.L. (1983). *Elements of generalizability theory*. Iowa City, IA: American College Testing Program.

Brown, F.G. (1981). *Measuring classroom achievement*. New York: Holt, Rinehart and Winston.

Cronbach, L.J., Gleser, G., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.

Cunningham, G.K. (1986). *Educational and psychological measurement*. New York: Macmillan.

Frisbie, D.A., & Friedman, S.J. (1987). Test standards—some implications for the measurement curriculum. *Educational Measurement: Issues and Practice, 6*(3), 17-23.

Gullickson, A.R. (1984). Teacher perspectives of their instructional use of tests. *Journal of Educational Research, 77*(4), 244-248.

Gullickson, A.R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. *Journal of Educational Measurement, 23*(4), 347-354.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer-Nijhoff.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mayo, S.T. (1964). What experts think teachers ought to know about educational measurement. *Journal of Educational Measurement, 1*(1), 79-86.

McLean, L.D. (1985). *The craft of student evaluation in Canada*. Toronto: Canadian Education Association.

Rudman, H.E., Kelley, J.L., Wanous, D.S., Mehrens, W.A., Clark, C.M. & Porter, A.C. (1980). *Integrating assessment with instruction: A review (1922-1980)* (Research Series No. 75). East Lansing: Michigan State University, Institute for Research on Teaching. (ERIC Document Reproduction Services No. ED 189 136)

Stiggins, R.J., & Bridgeford, N.J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271-286.

Stiggins, R., Conklin, N.F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice, 5*(2), 5-17.

Thorndike, R.L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

Wilson, R.J. & Rees, R. (in press). The ecology of evaluation: Measurement and evaluation within an administrative framework. *Canadian Journal of Education*.

The Alberta Journal of Educational Research *Vol. XXXV, No. 2, June 1989, 145-163*

L.L. MORTON
University of Windsor

P.H. LINDSAY
Ontario Institute for Studies in Education, University of Toronto

and

W.M. ROCHE
Northumberland and Newcastle Board of Education, Cobourg, Ontario

Word Processing Effects on Writing Productivity and Revision at Elementary and Junior High School Levels

A six-month study of writing performance contrasted pencil-and-paper (P&P) with lab-based word processing (W-P) for regular grade (5 to 8) children (N=36). In terms of productivity, W-P increased (a) number of drafts, story units, and overall output for all students; and (b) final draft length for grades 5 and 6 students. In terms of revisions, W-P increased (a) spelling, paragraphing, and formatting revisions; (b) meaning-preserving revisions for grade 6 students; and (c) meaning-altering, deletion revisions for grade 8 students. Strong linear trend effects were evident indicating developmental improvement. Deviation from linearity was due to notably poorer performance for the lower grade students, particularly grade 5 students. Grade 5 students appear significantly less able to use W-P technology. Overall, the results present an argument for the use of W-P at the elementary and junior high school levels and indicate that a lab setting is workable.

The purpose of this study was to explore word processing effects on writing in normal-achieving children. The rationale for this focus emerged from (a) the call for answers to questions on word processing effects as a precursor to the tendering and testing of theoretical hypotheses (Logan & Padro, n.d.), (b) the call for research to supplement the sparse data which are presently available on word processing effects (Collier, 1983; Grabe, 1985; Maddux, 1985), and (c) the call for clarification of the effects of word processing on revision (Fitzgerald, 1987).

Larry Morton is an assistant professor at the University of Windsor, Faculty of Education, with interests in exceptional learners, computer applications in education, and environmental influences on information processing.

Peter Lindsay is an associate professor in the Department of Special Education at the Ontario Institute for Studies in Education, with interests in cognitive psychology, exceptional learners, and the potential of computer applications in education.

Mike Roche is a science department head with the Northumberland and Newcastle Board of Education, with interests in computer applications in education.

The availability of a word processor does not presage positive effects. Research on composing is less than encouraging. Programs that interact with the writer by asking for input to generate a poem may teach poetic structure (Owens, 1984) yet may detract from creativity and originality. Moreover, programs that prompt the writer by asking directive questions in response to "help" keys or time delays may actually retard the composing process (Woodruff, Bereiter, & Scardamalia, 1981-82). Furthermore, there is some evidence that children may not be predisposed to *compose* on a word processor; rather, they may prefer to compose with pencil and paper and *revise* on the word processor (Woodruff, Lindsay, Bryson, & Joram, 1986).

A second problem concerns the measurement of word processing effects. Qualitative effects are often evaluated merely by amount written or subjective judgments (Bradley, 1982; Kleiman & Humphrey, 1982). In addition, the reported research effects are often more descriptive than statistical (MacArthur & Shneiderman, 1986; Sapona, Lloyd, & Wissick, 1986). In effect, qualitative improvement is not always evident (Woodruff et al., 1981-82) or is not clearly evident (MacArthur & Shneiderman, 1986; Sapona et al., 1986). Thus it is essential to have measures which convincingly demonstrate quantitative and qualitative effects.

The third problem concerns target populations. Existing word processing research is not child oriented (Bridwell, Nancarrow, & Ross, 1984; Collier, 1983; Fitzgerald, 1987).

Thus there seem to be legitimate concerns about the existing research base (Collier, 1983; Grabe, 1985; Maddux, 1985). There is a need to address these concerns when exploring word processing effects with children. The present study uses a comprehensive within-subjects design with a thorough statistical analysis in an attempt to transcend the limitations and subjective views in word processing research.

Intuitively, word processing seems to have the potential to enhance the writing of both normal achieving and communication disordered children. Beyond such benefits as motivation and aesthetically appealing printed products, word processing allows all students adult-like writing experiences (Papert, 1980). It has been suggested that word processing may facilitate language development (Bradley, 1982) and cognitive development (Alexander, 1983). Moreover, the central craft components (Graves, 1983) of writing may be improved by word processing technology.

In fact, word processing may prove to be one of the most significant educational applications of the microcomputer. It eliminates typewriters, pens, scratch paper, erasers, cut-and-paste tools, and possibly even dictionaries; moreover, the simplification of storing, editing, and revising text leaves students free for higher order processing (Kleiman & Humphrey, 1982).

Higher order aspects of creative writing include *planning*, *composing*, and *editing-revising*. It is the simplification due to the *editing-revision process* which appears to have the potential to facilitate creative writing by allowing for a constructive focus on content and revision (Graves, 1983).

The tedium associated with error corrections, text modifications (changes, additions, deletions, etc.), and subsequent drafts is eliminated.

Furthermore, there would be a recursivity established between thinking and writing by means of the revision process (Alexander, 1983). The fact that the writer cyclically moves back and forth between thinking and writing, while revising, may benefit both cognitive and language processes.

Given the existing research deficiencies and the intuitive appeal of word processing, a study was designed to explore word processing effects with children. Developmental data were collected on the writing performance of normal achieving children (grades 5-8) in both pencil-and-paper and word processing settings. The focus of the study was limited to the editing-revision process (in view of the findings of Woodruff et al., 1986) rather than the planning or composing processes. Consistent with the logic of popular opinion on the potential value of word processing, it was predicted that (a) word processing would increase productivity in terms of drafts produced, story units produced, number of words written on final drafts, total word output and revisions made, and (b) proficiency would increase with grade level.

Method

Subjects

Ten subjects were randomly selected from each of four grades (5-8) and matched by teachers according to sex, age, and achievement to form two groups at each grade level. Two subjects at each of the grades 5 and 6 levels were lost due to attrition and incomplete data, leaving a sample of 36 students. The teachers used the Whole Language approach for the language arts program. Writing activities emerged from integrated thematic studies where topics could be suggested by the teacher or selected by the student. Even so, the writing program was not considered to be a significant variable because a within-subjects design was used with appropriate counterbalancing for order of word processing exposure.

Test scores on the Test of Written Language (TOWL, Hammill & Larsen, 1983) indicated that student writing proficiency was within the average range (85 to 115, standard score mean=104.5).

Procedure

Each student was required to use the microcomputer for writing for a period of three months (approximately 30 sessions at 30 to 40 minutes per session) and, further, to use paper-and-pencil only in the writing program for a comparable three month period involving the same number of 30 to 40 minute sessions. One group of five students at each grade level was randomly selected to do the paper-and-pencil sessions first, while the second group did the word processing sessions first. Groups were reversed at the end of the three-month period. Because the purpose of the study was to examine the revision process, first drafts of all creative writing were done in pencil-and-paper sessions under teacher direction. Subsequent drafts were done in either pencil-and-paper or word processing, depending on group assignment.

Five microcomputers (Commodore 64s) were set up in an empty classroom along with one 1526 dot matrix printer. One additional computer with a Panasonic printer was set up in another room for final printing. Students (grades 5 to 8) in groups of five were booked into the computer lab for 30-minute sessions daily. Teachers monitored student writing from their classrooms and only occasionally had the opportunity to monitor the lab. Lab assistance was provided by a cooperative education student on work-assignment from a secondary school or by a knowledgeable grade 8 student. All students were entering, saving, and printing text within their first two sessions using the menu driven word processor *Better Working Word Processor* (Spinnaker, 1985).

A standardized writing test (TOWL) was used for pretest and posttest comparisons. On subtests used from this test, students were required to (a) complete sentences using a cloze format which evaluated for Word Usage (the use of correct verb tense, auxiliary verbs, number, comparatives, etc.); (b) proofread and correct sentences for punctuation and capitalization which evaluates for style; and (c) to write a story in response to a pictorial stimulus which was evaluated for Vocabulary (words used with a length greater than six letters) and Thematic Maturity (based on 20 questions relating to the use of humor, personal names, dialogue, etc.). Story length was also tabulated and recorded. Technical data on the TOWL for standardization, reliability, and validity are reported to be satisfactory (Salvia & Ysseldyke, 1981). All reliability coefficients reported in the manual exceed .80. Moreover, where a subjective element is involved (Thematic Maturity) interscorer reliability is reported as .93. With respect to validity, the authors report adequate criterion-related and construct validity.

The classroom teachers—who were unaware of the specific objectives of the study—monitored the writing of all students and provided motivation and editorial feedback. However, the feedback during word processing was less immediate in view of the teacher-student distance factor. Students in the lab were required to be more independent. The students revised their text in the lab and presented their revised draft (i.e., a printout) to the teacher for further editorial comment. The process continued until both teacher and student were satisfied with the draft which then became the final draft. The teachers coded each draft as first, second, third, and so on and filed a copy for subsequent evaluation for developmental changes in quantity and quality.

To determine productivity (i.e., quantitative changes) in both pencil-and-paper and word processing settings, frequency counts were made for (a) number of drafts produced (i.e., number of revised printouts on a story which were presented to the teacher for evaluation and feedback); (b) number of words in each draft; (c) number of story units (i.e., "final publications") generated over the three-month period; and (d) total output of each group in each setting.

To evaluate revisions (i.e., qualitative changes) a taxonomy modeled after Faigley and Witte (1984) was developed (Table 1). Text was evaluated for Infralexical Changes and Lexical Changes. The Infralexical Changes

Table 1
Taxonomy of Revision Types

Infralexical Revisions	Lexical Revisions
Level of Change	*Level of Change*
A. First-Order Changes	A. First-Order Changes (meaning-preserving)
1. Spelling (01)	1. Additions (11)
2. Punctuation (04)	2. Deletions (12)
	3. Substitutions (13)
B. Second-Order Changes	B. Second-Order Changes (meaning-altering-gist-preserving)
1. Tense, Number, etc. (02)	1. Additions (21)
2. Abbreviations (03)	2. Deletions (22)
	3. Substitutions (23)
C. Third-Order Changes	C. Third-Order Changes
1. Paragraphing (05)	1. Additions (31)
2. Formatting (06)	2. Deletions (32)
	3. Substitutions (33)

Note: The numbers in brackets are cross-references to the taxonomy used by Faigley and Witte (1984).

were considered to be below the level of the "word" and were catalogued as First-Order Changes (spelling and punctuation), Second-Order Changes (tense/number and abbreviations), and Third-Order Changes (paragraphing and formatting).

The Lexical Changes were considered to be at the "word" level or above and involved additions, deletions, or substitutions. At the First-Order level the changes were considered to be meaning-preserving; at the Second-Order level the changes were defined as meaning-altering but gist-preserving; and at the Third-Order level the changes were considered to be meaning-altering and gist-altering. By way of explanation, a gist-alteration in a passage would result from a change in the body of the text (e.g., addition) that would also require a change in the written "abstract." A change in the body of the text which changed the meaning but did not require a change in the "abstract" would be a gist-preserving change—a Second-Order change.

To apply this taxonomy in a manageable way and within a standardized format, the first 30 words in each revised draft were evaluated for revisions and then tabulated. The rationale for a 30-word limit emanated from the

need to keep the data manageable and the observation that there were no written story units smaller than 30 words in length.

Pretest and posttest scores using standardized scores (i.e., standard scores) from four subtests (Word Usage, Style, Vocabulary, and Thematic Maturity) of the TOWL were evaluated for both pencil-and-paper and word processing conditions.

Results

The data were analyzed using the mixed-model analysis of variance (ANOVA) with repeated measures, followed by trend analysis (F-tests) and simple effects tests (t-tests) where appropriate. The results are presented in four sections: Writing Productivity, Lexical Revisions, Infralexical Revisions, and Test of Written Language.

Writing Productivity

Number of drafts. A two-way (4x2) ANOVA with Grade (5, 6, 7, 8) and Setting (word processing, pencil-and-paper) as the independent variables was computed for number of drafts produced. Setting was a within-subjects variable treated as a repeated measure. There was a main effect for Setting, $F(1,32)=30.40$, $p<.0001$, with the word processing grouping producing more drafts (mean=2.54) than the paper-and-pencil grouping (mean=1.78). Also, there was a main effect for Grade, $F(3,32)=40.13$, $p<.0001$, with no interaction effects. As may be seen in Figure 1a, however, there was a strong linear trend, $F(1,32)=54.24$, $p<.01$, with an increase in drafts at the higher grade levels. Deviation from linearity, $F(2,32)=33.0$, $p<.01$, was apparently due to lower performance overall at the lower grade levels and a plateau at the higher grades for pencil-and-paper. Simple effects tests confirmed (a) that the grades 5 and 6 students were performing lower than the grades 7 and 8 students, $p<.01$; and (b) only the grade 5 students did not show a significant difference between word processing and paper-and-pencil settings, $p>.05$.

Number of story units. A two-way (4x2) ANOVA with Grade and Setting as the independent variables was computed for number of story units produced. There was a main effect for Setting, $F(1,32)=5.68$, $p<.05$, with the word processing grouping producing more story units (mean=11.92) than the paper-and-pencil grouping (mean=10.1), and a main effect for Grade, $F(3,32)=4.98$, $p<.01$. The significant linear trend for Grade, $F(1,32)=6.62$, $p<.05$, was affected by significant deviation, $F(2,32)=4.16$, $p<.05$. Tests for simple effects revealed lower performance in story unit production for grades 5 and 6 students (Figure 1b) when compared to the grades 7 and 8 students but only during the paper-and-pencil sessions, $p<.05$. Thus word processing seems to increase story unit production for lower grade students.

Total words produced. For total output a two-way (4x2) ANOVA (Grade by Setting) was computed for number of words produced. There was a main effect for Setting, $F(1,32)=24.28$, $p<.0001$, with the word processing grouping producing more words (mean=4560.75) than the paper-and-pencil grouping (mean=2711.33). Also, there was a main effect for Grade,

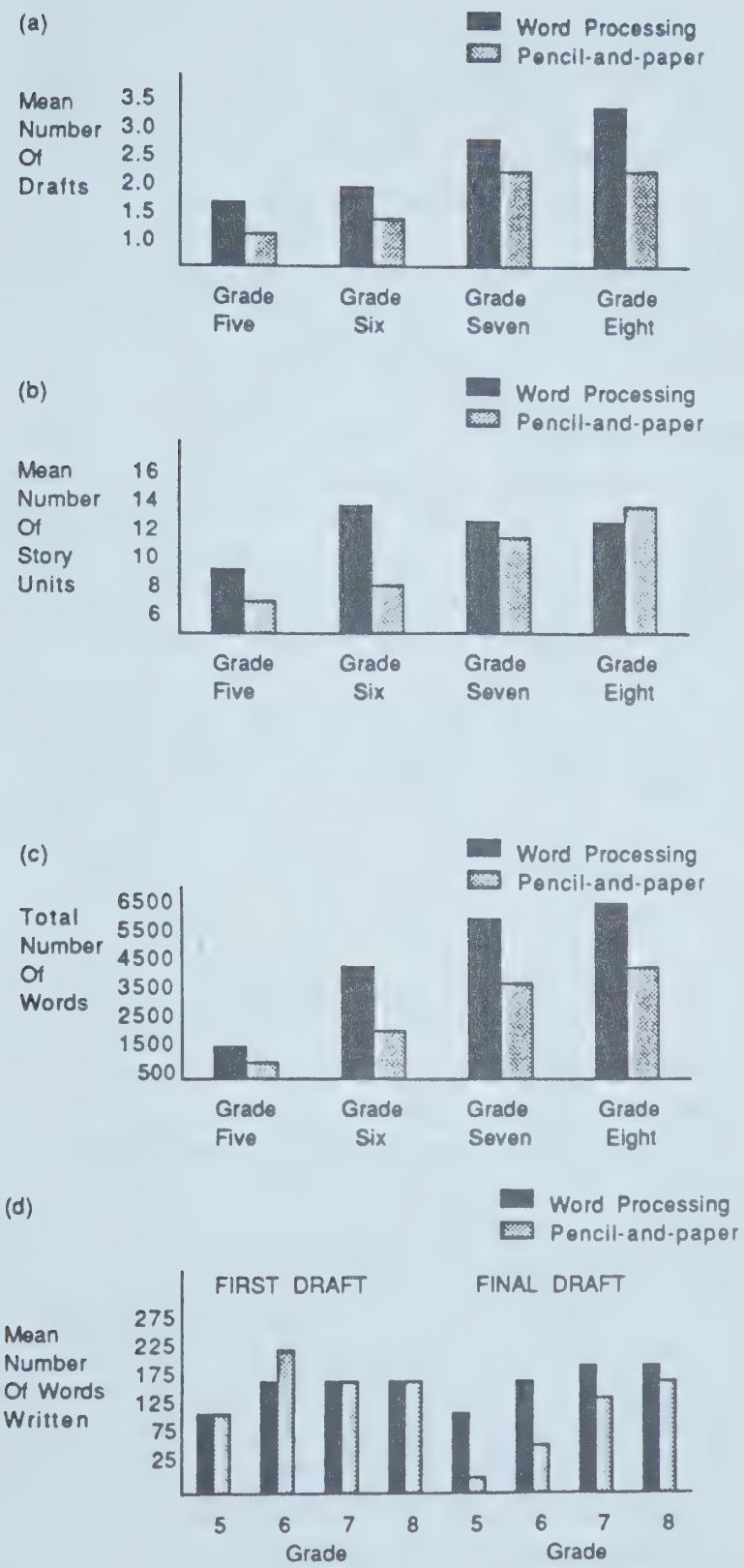


Figure 1: Written productivity for students in grades 5 through 8 during word processing and paper-and-pencil sessions in terms of (a) number of drafts produced, (b) number of story units generated, (c) total number of words produced, and (d) number of words written for first and final drafts

$F(3,32)=18.44$, $p<.0001$. The significant linear trend for Grade, $F(1,32)=26.27$, $p<.01$, was affected by significant deviation, $F(2,32)=14.51$, $p<.01$. In the paper-and-pencil sessions, simple effects tests revealed lower performance by grades 5 and 6 students when compared to the grades 7 and 8 students, $p<.01$. In the word processing sessions all students appeared to benefit (Figure 1c); however, the difference for grade 5 students between pencil-and-paper and word processing was not significant, $p>.05$.

Draft lengths. A three-way (4x2x2) ANOVA with Grade, Setting, and Draft as the independent variables was computed for length of first and final drafts. There was a main effect for Grade, $F(3,32)=4.39$, $p<.01$, which was qualified by a two-way interaction for Grade and Draft, $F(3,32)=3.47$, $p<.05$, and a two-way interaction for Setting and Draft, $F(1,32)=7.21$, $p=.01$. As may be seen in Figure 1d, the grades 7 and 8 students showed increases in final draft length, whereas the grades 5 and 6 students showed decreases. Tests for simple effects revealed that the grade 8 students showed final draft increases in both word processing, $p<.04$, and pencil-and-paper settings, $p<.05$, whereas the grade 7 students showed increases in the word processing sessions only, $p<.03$. Furthermore, the grade 5 students showed a dramatic absence of final draft length during pencil-and-paper sessions, $p<.005$, but not during word processing, $p>.05$. A linear trend was not evident for overall draft length, $F(1,32)= 2.1$, $p>.05$. Means and standard deviations for number of drafts, number of story units, number of words written, and length of first and final drafts are reported in Table 2.

Lexical Revisions

A four-way (4x2x3x3) ANOVA with Grade (5, 6, 7, 8), Setting (word processing, pencil-and-paper), Level of Change (First-Order, Second-Order, Third-Order), and Revision Type (addition, deletion, substitution) as the independent variables was computed. The final three variables were within-subjects variables treated as repeated measures. There were main effects for Grade, $F(3,32)=34.96$, $p<.0001$; Level of Change, $F(2,64)=42.14$, $p<.0001$; and Revision Type, $F(2,64)=20.64$, $p<.0001$; and two-way interactions for Grade and Setting, $F(3,32)=3.87$, $p<.02$; Grade and Level of Change, $F(6,64)=10.37$, $p<.0001$; and Grade and Revision Type, $F(2,64)=5.95$, $p<.0001$. In addition, there were three-way interactions for Grade, Setting, and Revision Type, $F(6,64)=3.83$, $p<.003$; and Grade, Level of Change, and Revision Type, $F(12,128)=2.08$, $p<.02$. To clarify these interactions, each Level of Change was analyzed separately. The descriptive data on these separate effects are presented graphically in Figure 2.

First-order lexical revisions. A 4x2x3 ANOVA (Grade by Setting by Revision Type) on the First-Order revisions revealed main effects for Grade, $F(3,32)=13.55$, $p<.0001$; and Revision Type, $F(2,64)=5.95$, $p<.005$; and two-way interactions for Grade and Setting, $F(3,32)=5.78$, $p<.003$; and Grade and Revision Type, $F(6,64)= 3.17$, $p<.01$. The linear trend showing increased revisions at the higher grade levels was significant, $F(1,32)=6.37$, $p<.05$, as was the deviation from linearity, $F(2,32)= 17.14$, $p<.01$. Tests for

Table 2
Means and Standard Deviations for Word Processing and Pencil-and-Paper Sessions at Four Grade Levels*

	Grade 5	Grade 6	Grade 7	Grade 8
Number of Drafts				
W-P	1.8 (1.0)	1.9 (0.3)	3.0 (0.5)	3.2 (0.6)
P&P	1.1 (0.3)	1.3 (0.4)	2.2 (0.3)	2.2 (0.4)
Number of Stories				
W-P	9.1 (4.7)	13.6 (4.1)	12.4 (2.0)	12.3 (3.6)
P&P	7.0 (3.6)	7.8 (4.3)	12.0 (2.8)	12.7 (3.9)
Number of Words				
W-P	1623 (1039)	4082 (1381)	5580 (1964)	6275 (2528)
P&P	833 (616)	1718 (1485)	3708 (1385)	4012 (1699)
First Draft Length				
W-P	107.4 (38.0)	163.5 (86.7)	143.1 (57.8)	148.0 (47.7)
P&P	101.5 (26.6)	218.9 (279.5)	139.1 (43.0)	131.6 (50.0)
Final Draft Length				
W-P	94.8 (47.3)	154.7 (43.5)	168.3 (66.0)	173.8 (66.8)
P&P	9.4 (26.6)	63.3 (69.6)	149.1 (52.8)	162.3 (85.5)

Note: W-P = Word Processing and P&P = Pencil-and-Paper.
*Standard Deviations in brackets

simple effects revealed no differences in the revision incidence for grades 6, 7, and 8 in the word processing setting; however, the grade 5 students showed a lower incidence of revision on all three revision types, $p<.05$. In the pencil-and-paper sessions there were no differences between the grades 5 and 6 students nor between the grades 7 and 8 students. However, the grades 5 and 6 students showed a lower incidence of revision than the grades 7 and 8 students, $p<.01$. Thus it was the grade 6 students who seemed to be more able to take advantage of the word processing setting when compared with their pencil-and-paper performance.

Second-order lexical revisions. A 4x2x3 ANOVA (Grade by Setting by Revision Type) on the Second-Order revisions revealed main effects for Grade, $F(3,32)=37.08$, $p<.0001$; and Revision Type, $F(2,64)=13.11$, $p<.0001$; a two-way interaction for Grade and Revision Type, $F(6,64)=3.47$, $p<.01$; and a three-way interaction for Grade, Setting, and Revision Type, $F(6,64)=3.57$, $p<.005$. Linear trend was not significant ($F<1$). Tests for simple effects revealed lower revision incidence for grades 5 and 6 students when compared to grades 7 and 8 students in the word processing setting, $p<.05$, and the pencil-and-paper setting, $p<.01$ (Figure 2b). When

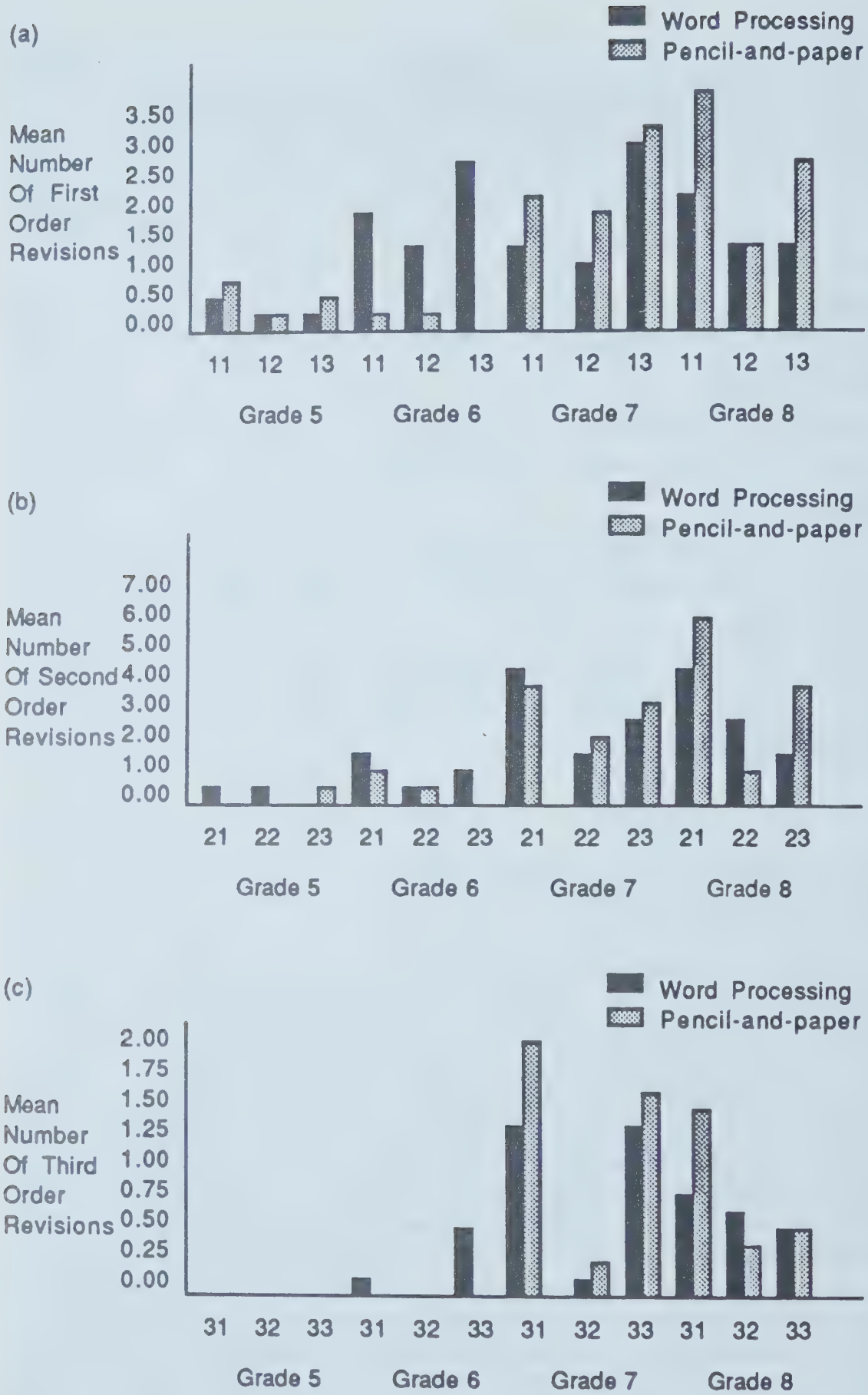


Figure 2: Mean number of lexical revisions (additions, deletions, and substitutions) at the First-Order level (11, 12, 13), the Second-Order level (21, 22, 23), and the Third-Order level (31, 32, 33) for students in grades 5 through 8 in word processing and pencil-and-paper sessions in the first 30 words of revised drafts

grade level was considered separately, only grade 8 students showed a difference between word processing and pencil-and-paper, and there only for deletion revisions which increased during the word processing sessions, $p<.01$.

Third-order lexical revisions. A 4x2x3 ANOVA (Grade by Setting by Revision Type) on the Third-Order revisions revealed main effects for Grade, $F(3,32)=13.50$, $p<.0001$, and Revision Type, $F(2,64)=7.94$, $p<.001$, which were qualified by a two-way interaction for Grade and Revision Type, $F(6,64)=4.30$, $p<.01$. Linear trend was significant $F(1,32)=4.21$, $p<.05$, as was deviation from linearity, $F(2,32)=18.14$, $p<.01$. In terms of grade level the grades 7 and 8 students showed superior revision incidence for addition revisions only, $p<.01$, in both the word processing and pencil-and-paper settings (Figure 2c). Means and standard deviations are reported in Table 3.

Infralexical Revisions

At the Infralexical level, a four-way (4x2x3x2) ANOVA (Grade by Setting by Level of Change by Revision Type) was computed. There were main effects for Grade, $F(3,32) =11.61$, $p<.0001$; Setting, $F(3,32)=21.93$, $p<.0001$; Level of Change, $F(2,64)=50.07$, $p<.0001$; and Revision Type, $F(1,32)=1.21$, $p<.005$. Two-way interactions were evident for Grade and Level of Change, $F(6,64)=3.11$, $p<.01$; Setting and Level of Change, $F(2,64)=8.58$, $p<.001$; and Setting and Revision Type, $F(1,32)=6.26$, $p<.02$. Also, there was a three-way interaction for Setting, Level of Change, and Revision Type, $F(2,64)=12.50$, $p<.0001$, and a four-way interaction for Grade, Setting, Level of Change, and Revision Type, $F(6,64)=2.87$, $p<.02$. To clarify these effects, each Level of Change was analyzed separately. The descriptive data on these separate effects are presented graphically in Figure 3.

First-order infralexical revisions. A 4x2x2 ANOVA (Grade by Setting by Revision Type) on the First-Order revisions (spelling, punctuation) revealed main effects for Grade, $F(3, 32)=5.79$, $p<.003$; Setting, $F(1,32)=11.39$, $p<.003$; and Revision Type, $F(1,32)=6.70$, $p<.02$. There was a two-way interaction for Setting and Revision Type, $F(1,32)=15.75$, $p<.001$, and a three-way interaction for Grade, Setting, and Revision Type, $F(3, 32)=3.11$, $p<.05$. Linear trend was significant, $F(1,32)=4.30$, $p<.05$, as was the deviation from linearity, $F(2,32)=6.52$, $p<.01$. Simple effects tests on the revision type revealed that the higher revision incidence in word processing was due to the increase in spelling revisions, $p<.001$ (Figure 3a), although this was not true for grade 5 students, $p>.05$. Again, the grade 5 students seemed less able to effectively use the word processor beyond the paper-and-pencil level.

Second-order infralexical revisions. A similar three-way (4x2x2) ANOVA (Grade by Setting by Revision Type) on the Second-Order revisions (tense/number, abbreviation) revealed main effects for Grade, $F(3,32)=8.32$, $p<.001$; and Revision Type, $F(1,32)=25.15$, $p<.0001$. Also, there was a two-way interaction for Grade and Revision Type, $F(3,32)=7.30$, $p<.001$. Linear trend for grade level was significant, $F(1,32)=5.54$, $p<.05$, as was the deviation from linearity, $F(2,32)=9.70$, $p<.01$. Simple effects tests on

Table 3
Lexical Revisions with Word Processing and Pencil-and-Paper

		Word Processing Grade				Pencil-and-Paper Grade			
		5	6	7	8	5	6	7	8
First Order									
Additions	M	0.25	1.75	1.10	2.00	0.63	0.13	1.90	3.50
	SD	0.46	1.75	1.20	1.49	1.77	0.35	1.20	2.37
Deletions	M	0.13	1.12	1.00	1.18	0.13	0.13	1.50	1.20
	SD	0.35	1.13	1.15	1.87	0.35	0.35	0.97	1.55
Substitutions	M	0.13	2.63	2.26	1.20	0.38	0.00	3.00	2.50
	SD	0.35	2.72	1.83	1.48	1.06	0.00	2.00	1.58
Second Order									
Additions	M	0.13	1.13	3.90	4.20	0.00	0.25	3.50	5.50
	SD	0.35	1.46	3.51	2.70	0.00	0.46	1.58	2.80
Deletions	M	0.13	0.13	1.20	2.10	0.00	0.13	1.80	0.90
	SD	0.35	0.35	1.40	0.88	0.00	0.35	1.48	1.10
Substitutions	M	0.00	0.75	2.10	1.40	0.13	0.00	3.00	3.60
	SD	0.00	1.16	1.37	1.51	0.35	0.00	2.45	2.84
Third Order									
Additions	M	0.00	0.13	1.30	0.80	0.00	0.00	2.00	1.40
	SD	0.00	0.35	1.06	0.92	0.00	0.00	1.83	1.90
Deletions	M	0.00	0.00	0.10	0.70	0.00	0.00	0.20	0.40
	SD	0.00	0.00	0.32	0.82	0.00	0.00	0.42	0.52
Substitutions	M	0.00	0.38	1.30	0.60	0.00	0.00	1.50	0.60
	SD	0.00	1.06	1.34	0.97	0.00	0.00	1.51	0.84

the revision types revealed a higher incidence of tense/number revisions, $p<.01$; however, this was due to the grades 7 and 8 students, $p<.05$, not the grades 5 and 6 students (Figure 3b). Because setting was not significant, word processing does not seem to affect Second-Order Infralexical revisions at the grade levels studied.

Third-order infralexical revisions. A 4x2x2 ANOVA (Grade by Setting by Revision Type) on the Third-Order revisions (paragraphing and formatting) revealed main effects for Grade, $F(3,32)=3.47$, $p<.03$, and Setting, $F(1,32)=44.00$, $p<.0001$. As may be seen in Figure 3c, word processing clearly facilitates Third-Order revisions. There were no interactions. Linear

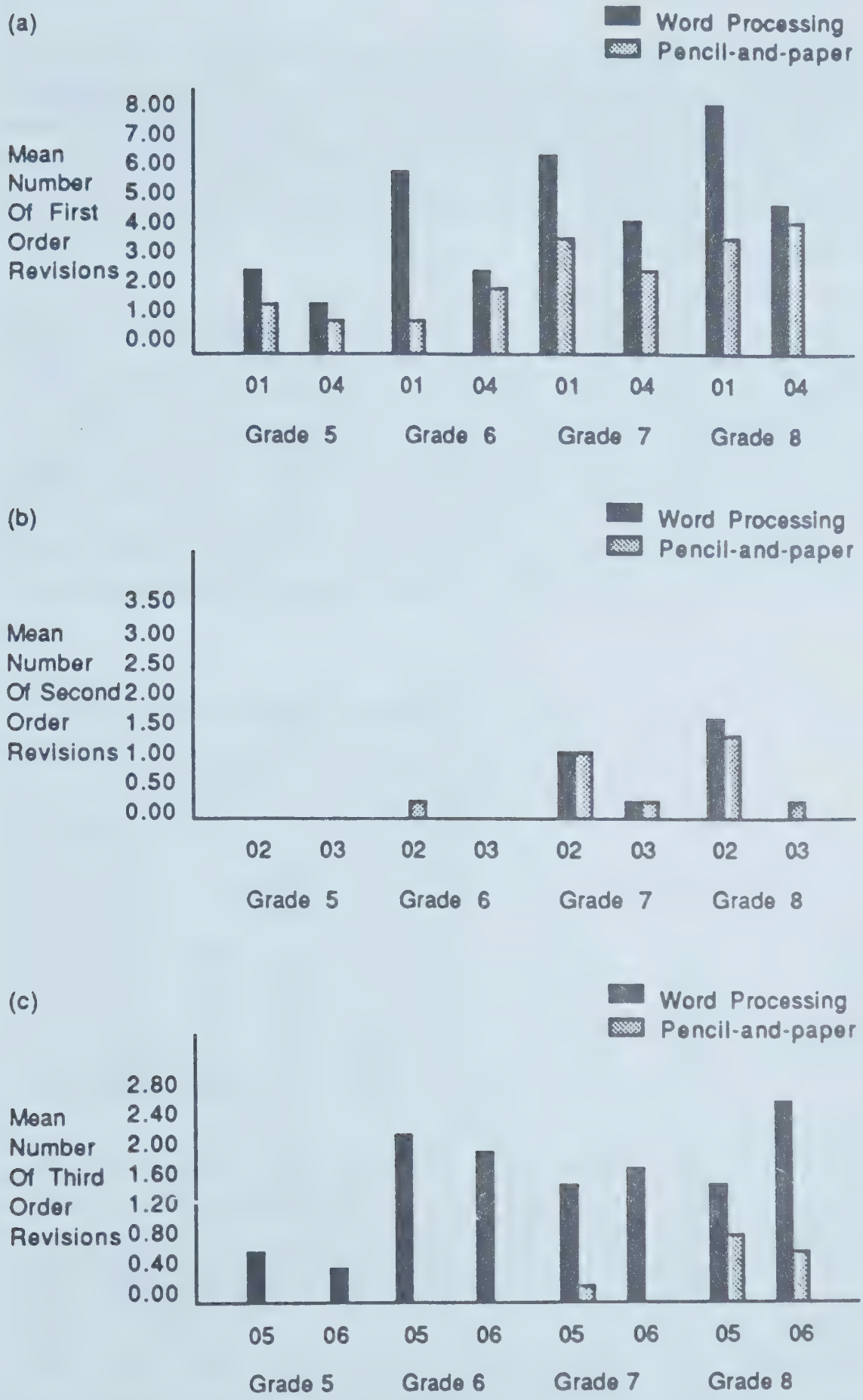


Figure 3: Mean number of infralexical revisions for students in grades 5 through 8 in word processing and pencil-and-paper sessions in the first 30 words of revised drafts. First-Order level: spelling (01), punctuation (04); Second-Order level: tenses (02), abbreviations (03); Third-Order level: paragraphing (05), formatting (06)

trend for grade level was not significant. Simple effects tests on the revision types revealed that the lower incidence of revision in word processing for grade 5 students differed significantly only from the grade 8 students, $p<.05$. However, in view of the general equivalence in performance between grades 6, 7, and 8 students (Figure 3c), the main effect for Grade would appear to be due to the poor grade 5 performance. Means and standard deviations are presented in Table 4.

Test of Written Language

Standard scores. A four-way (4x2x2x4) ANOVA on the TOWL standard scores (Table 5) with Grade (5, 6, 7, 8), Setting (word processing, pencil-and-paper), Test Time (pretest, posttest), and Subtest (Vocabulary, Thematic Maturity, Word Usage, Style) as the independent variables, with the last three variables being within-subjects variables treated as repeated measures, revealed no significant effects for Setting. A main effect for Test Time, $F(1,32)=15.54$, $p<.001$, indicated higher posttest scores in both settings. A main effect for Subtest, $F(3,96)=7.21$, $p<.001$, reflects generally lowered performance on the Style Subtest relative to the norming group.

Table 4
Infralexical Revisions with Word Processing and Pencil-and-Paper

		Word Processing Grade				Pencil-and-Paper Grade			
		5	6	7	8	5	6	7	8
First Order									
Spelling	M	2.50	5.38	5.80	8.20	1.25	0.75	3.60	3.20
	SD	2.39	4.17	3.58	5.43	3.54	0.89	3.40	2.62
Punctuation	M	1.25	2.00	4.20	4.80	0.50	1.63	2.70	4.40
	SD	1.90	2.00	2.20	4.02	1.41	4.60	2.00	2.50
Second Order									
Tense, etc.	M	0.00	0.00	1.10	1.50	0.00	0.25	1.10	1.30
	SD	0.00	0.00	1.37	1.78	0.00	0.71	1.10	1.16
Abbreviations	M	0.00	0.00	0.10	0.00	0.00	0.00	0.20	0.10
	SD	0.00	0.00	0.32	0.00	0.00	0.00	0.42	0.32
Third Order									
Paragraphing	M	0.50	2.00	1.40	1.30	0.00	0.00	0.25	0.70
	SD	0.53	2.27	2.01	1.06	0.00	0.00	0.32	1.06
Formatting	M	0.38	1.75	1.60	2.50	0.00	0.00	0.00	0.50
	SD	0.74	1.75	1.65	2.27	0.00	0.00	0.00	0.71

Table 5
Pretest and Posttest Standard Scores on the Test of Written Language (TOWL) for Word Processing and Pencil-and-Paper Conditions

		Word Processing Grade				Pencil-and-Paper Grade			
		5	6	7	8	5	6	7	8
Pretest									
Vocabulary	M	11.3	10.5	10.8	11.5	11.1	10.5	10.2	11.8
	SD	3.0	3.1	2.5	2.9	2.9	2.0	3.1	2.3
Thematic	M	10.6	11.5	10.2	10.8	10.8	10.9	9.9	10.4
Maturity	SD	2.3	1.7	2.3	2.0	2.0	1.4	2.1	2.0
Word Usage	M	11.4	12.5	11.1	10.9	10.3	12.3	10.0	11.4
	SD	2.8	1.6	2.1	2.7	1.5	1.8	1.6	2.5
Style	M	9.5	10.5	8.6	9.7	9.6	10.3	8.4	10.2
	SD	0.9	1.1	3.0	2.8	1.3	1.3	3.0	2.7
Posttest									
Vocabulary	M	10.6	10.8	10.7	12.0	10.6	11.3	11.0	11.4
	SD	3.0	3.1	2.5	2.9	3.3	3.8	2.7	2.7
Thematic	M	10.6	11.8	10.3	10.9	10.5	12.0	10.5	11.3
Maturity	SD	3.0	3.1	2.5	2.9	2.3	2.2	2.3	2.3
Word Usage	M	11.8	13.6	11.0	11.4	11.8	13.1	11.1	11.2
	SD	3.0	3.1	2.5	2.9	2.7	1.6	1.9	2.7
Style	M	9.8	10.5	8.9	9.7	9.9	11.3	9.5	9.6
	SD	3.0	3.1	2.5	2.9	2.0	1.7	3.0	2.9

Note: The standard scores have a mean of 10 and a standard deviation of 3.

There were no interactions. Thus revision mode does not influence written language test performance.

TOWL story lengths. A three-way (4x2x2) ANOVA (Grade by Setting by Test Time) for the number of words written on the TOWL pretests and posttests was nonsignificant, thus indicating all groups were writing stories of comparable length both before and after the word processing and pencil-and-paper sessions.

Discussion

The prediction of increased productivity with a word processor was supported. When students were assigned to the word processing session they produced (a) more draft copies of revised texts, (b) more story units, and (c) more words in total output. Furthermore, the word processing setting

facilitated the production of final drafts, especially for grade 5 students. This is consistent with previous research indicating increased final draft length following revision (Morton, in press; Woodruff et al., 1986). Thus word processing is seen to be of value in producing quantitative changes in writing.

The qualitative changes were more localized. On Lexical Revisions, effects were noted which seemed to indicate developmental factors. First, word processing facilitated the meaning-preserving revisions for the grade 6 students in particular. As may be seen in Figure 2a, grades 5 and 6 students showed an absence of meaning-preserving (First-Order) revisions during pencil-and-paper sessions. However, during word processing sessions the grade 6 students showed an increased revision rate comparable to grades 7 and 8 students. Apparently, the grade 6 students have the ability to use meaning-preserving revisions, but the paper-and-pencil setting is an interfering rather than a facilitating factor. The grade 5 students appear to lack this proficiency regardless of mode of output. Conversely, the grades 7 and 8 students manifest proficiency in either mode of output. Thus the grade 6 students show particular qualitative benefits from the use of the word processor.

Second, the grade 8 students showed an increase in meaning-changing deletion revisions during the word processing sessions. This may be especially significant in that text deletion is a low frequency revision characteristic. Graves (1983) claims that, "it is a long time before any writer spontaneously wants to delete information" (p. 159). He further points out that, developmentally, text deletion occurs with teacher help prior to independent deletion practices. Yet in the present study, text deletion was seen to be contingent on the word processing setting. Thus this sophisticated aspect of word processing may reflect a developmental maturity factor for grade 8 students which is allowed to surface independent of teacher direction because of the word processing technology.

On Infralexical Revisions, word processing led to (a) an increase of spelling revisions, although this was not true for grade 5 students; and (b) a general increase of both paragraphing and formatting revisions. Thus, at both the Infralexical level and the Lexical levels, word processing is seen to benefit writing.

The prediction of developmental progression was also supported. Increasing linear trend over grade levels was generally evident. The deviation from linearity appeared to be due to several clear grade level differences.

Generally the grades 5 and 6 students were less proficient than the grades 7 and 8 students. Moreover, intragrade analysis showed that grade 5 students were not able to use the word processor to exceed paper-and-pencil performance levels. This may be an artifact of curricular differences, or it may reflect a discrete developmental difference in acquiring and using word processing skills. Grade 5 students in particular may be too young to take advantage of lab-based word processing.

However, at times the lower grade level students responded favorably to the word processing sessions (e.g., by increasing the final draft length). This

indicates a potentially positive effect for word processing. Similarly, the grade 6 students generated more story units during word processing. However, it is not clear whether this is due to motivation or to the word processing technology.

With respect to revisions, developmental effects were evident also. Generally the grades 5 and 6 students showed reduced revision proficiency. The grade 5 students were notably lacking on infralexical revisions—spelling, paragraphing, and formatting. Again, this would seem to suggest that the grade 5 students are more limited in effectively using the word processor in a lab setting.

Overall, however, word processing appears to facilitate the productivity and revisional aspects of writing. This finding adds to the existing research on revision (Fitzgerald, 1987) by showing not only that word processing has positive effects, but also that these positive effects extend, in some ways, to young children. Older students may be using the technology more effectively, but all students were seen to benefit from the word processing setting. Introducing systematic word processing practices to elementary school settings, therefore, would seem to herald positive effects.

However, the current lack of computers presents a problem (Maddux, 1985; Zaharias, 1983). Maddux (1985) notes that elementary students average about 30 minutes per week of computer time. He suggests that "schools may be adopting a policy of giving very light exposure to a large number of students, rather than intensive exposure to a smaller number" (p. 29). Such a time distribution would not be conducive to word processing. However, in the present study an intensive exposure was provided successfully to small groups of students.

The computer lab may be an ideal way to intensify exposure. By slotting time in a lab, the entire school may be able to increase access to computers. The process would be more systematized and computer "down time" would be minimal. Simple arithmetic suggests that 20-minute time slots in a lab with five computers could accommodate 70 students per day. If students had access to a time slot every day, or every other day, small schools could probably meet the needs of upper grade level (grades 6, 7, and 8) students. Larger schools would need larger or multiple labs.

Ideally, a teacher-run lab would alleviate monitoring problems and allow for effective interaction and conferencing. In fact, Morocco and Neuman (1986) see the main advantage of word processing in the access the teacher has to the process the child is using in writing. The text is more public, thus allowing for a more interactive process with other students and the teacher.

Unfortunately, the expense of a teacher-run lab makes it unlikely at the elementary level. However, computer labs could be attached to a library setting (with a computer-literate librarian) or to remedial and special education settings. Also, peer monitors, cooperative education students, and parent volunteers may work well in a lab setting as mediators between teachers and students.

The effects in this study are encouraging and serve to justify further explorations and innovations that would support word processing at the

elementary school level. For example, future research could examine the interface between specific writing programs and word processing technology. Some approaches to writing might be more amenable to the use of a word processor. Another possibility relates to the viability of using a lab setting with communication disordered students. Word processing technology may be a valuable tool in helping learning disordered students who have difficulty with writing. Also, a more stringent analysis of the cognitive processes involved during word processing would be an appropriate pursuit. This could be accomplished on the one hand by introducing "think-aloud" protocols and on the other hand by shifting the focus to include the planning and composing stages of writing on a word processor. In fact, a great deal of research remains to be undertaken in the field of computer applications (Logan & Padro, n.d.).

Notes

1. This research was supported by a grant from the Ontario Public School Teachers' Federation and by the Program Department of the Northumberland and Newcastle Board of Education.
2. The efforts and assistance of the principal and teachers—Gary Weekes, John Hayden, Judy O'Leary, Art Joynt, and Jim Dove—involved in this study are gratefully acknowledged.

References

Alexander, N. (1983). A primer on developing a writing curriculum. *Topics in Learning and Learning Disabilities, 3*(3), 55-62.

Bradley, V.N. (1982). Improving students' writing with microcomputers. *Language Arts, 59*, 732-743.

Bridwell, L.S., Nancarrow, P.R., & Ross, D. (1984). The writing process and the writing machine: Current research on word processors relevant to the teaching of composition. In R. Beach & L. S. Bridwell (Eds.), *New directions in composition research* (pp. 381-398). New York: Guilford Press.

Collier, R.M. (1983). The word processor and revision strategies. *College Composition and Communication, 34*, 149-155.

Faigley, L., & Witte, S.P. (1984). Measuring the effects of revision on text structure. In R. Beach & L. S. Bridwell (Eds.), *New directions in composition research* (pp. 95-108). New York: Guilford Press.

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 57*, 481-506.

Grabe, M. (1985). Evaluating the educational value of microcomputers. *Computers in the Schools, 2*, 35-44.

Graves, D.H. (1983). *Writing: Teachers and children at work*. Exeter: Heinemann Educational Books.

Hammill, D.D., & Larsen, S.C. (1983). *The test of written language*. Austin, TX: Pro-Ed.

Kleiman, G., & Humphrey, M. (1982, March). Word processing in the classroom. *Compute, 22*, 96-99.

Logan, R.K., & Padro, S. (n.d.). *Research questions on the impact of computers in the classroom*. Education and Technology Series. Ontario: Ministry of Education. (Contract 0601)

MacArthur, C.A., & Shneiderman, B. (1986). Learning disabled students' difficulties in learning to use a word processor: Implications for instruction and software evaluation. *Journal of Learning Disabilities, 19*, 248-253.

Maddux, C.D. (1985). Research paradigms: The need for breadth as well as depth. *Computers in the Schools, 2*, 27-30.

Morocco, C.C., & Neuman, S.B. (1986). Word processors and the acquisition of writing strategies. *Journal of Learning Disabilities, 19*, 243-247.

Morton, L.L. (in press). Word processing and the editing-revising process. *Computers in the Schools.*

Owens, P. (1984). Creative writing with computers. *Popular Computing, 3*(3), 128-132.

Papert, S. (1980). *Mindstorms*. New York: Basic Books.

Salvia, J., & Ysseldyke, J.E. (1981). *Assessment in special and remedial education*. Boston: Houghton Mifflin.

Sapona, R.H., Lloyd, J.W., & Wissick, C.A. (1986). Microcomputer use in resource rooms with learning disabled children. *Computers in the Schools, 2*, 51-60.

Spinnaker. (1985). *Better working word processor*. Spinnaker Software Corporation.

Woodruff, E., Bereiter, C., & Scardamalia, M. (1981-82). On the road to computer assisted compositions. *Journal of Educational Technology Systems, 10*, 133-148.

Woodruff, E., Lindsay, P., Bryson, M., & Joram, E. (1986, April). *Some cognitive effects of word processors on enriched and average eighth grade writers*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Zaharias, J.A. (1983). Microcomputers in the language arts classroom: Promises and pitfalls. *Language Arts, 60*, 990-996.

The Alberta Journal of Educational Research *Vol. XXXV, No. 2, June 1989, 164-170*

COLLEEN S. HAMMERMASTER
University of Alberta

Levels of Performance and Cognitive Interference in Test-Anxious Subjects

The major purpose of this study was to determine whether highly text-anxious subjects, in situations that pose testlike challenges, perform at relatively low levels and experience relatively high levels of task-irrelevant thoughts. The Test Anxiety Scale (TAS) was used to distinguish high and low test-anxious subjects. The subjects were individually administered the Wisconsin Card Sorting Test (WCST) as a measure of cognitive performance. Immediately after the WCST, subjects were asked to complete the Cognitive Interference Questionnaire (CIQ) which assesses irrelevant thinking that interferes with concentration on a task. Significant differences were found between the cognitive performance of high and low test-anxious groups, with high-anxious subjects performing less adequately. The high test-anxious subjects also experienced significantly more cognitive interference or task-irrelevant thinking than did their less anxious counterparts. A significant positive relationship between level of test-anxiety and amount of cognitive interference experienced was established.

According to the theoretical position adopted for this research, worry and other task-irrelevant cognitive activities associated with anxiety impair the quality of task-relevant performance. One explanation for the deterioration in performance is that the task-irrelevant information involved in worry and cognitive self-concern competes with task-relevant information for space in the processing system. As a consequence of critical self-focusing, highly test-anxious subjects are in effect in a dual-task or divided attention situation, in contrast to nontest-anxious subjects who primarily process task-relevant information.

This study compares levels of negative, self-deprecatory thoughts in high and low anxious groups and further explores the impact of this task-irrelevant interference on performance. Unlike many test-anxiety studies that have measured performance in terms of overall success or failure, this study examines particular components of learning that may inhibit or facilitate overall task performance. Inefficient initial conceptualization, perseveration, conceptual-level responses, failure to maintain set, and inefficient learning across several stages of a test are all evaluated as pos-

---

Colleen Hammermaster, a doctoral student in educational psychology, is currently involved in research examining the memory deterioration of persons with dementia of the Alzheimer type.

sible factors involved in the anticipated performance decrements of highly test-anxious persons.

Deffenbacher (1978) investigated sources of interference in highly test-anxious subjects who were solving difficult anagrams. Major findings were that, under evaluative stress, the high-anxiety group: (a) reported more anxiety during testing; (b) rated themselves, their abilities, and the task more negatively; (c) solved fewer anagrams; (d) experienced more interference from anxiety; and (e) reported greater distraction of attention due to heightened autonomic arousal (emotionality), worrisome thoughts (worry), and task-produced competing responses (task-generated interference) than did low-anxiety individuals under evaluative stress.

In a review of the literature on paper-and-pencil anxiety scales (Sarason, 1960), a number of studies were cited that provide evidence that highly test-anxious subjects are more "deprecatory, more self-preoccupied, and generally less content with themselves than subjects lower in the distribution of test anxiety scales" (p. 404). In addition to studies based on paper and pencil measures of anxiety, other studies indicate the self-deprecatory, self-ruminative tendencies of highly test-anxious subjects. In one such study (Sarason, 1960), highly test-anxious and low anxious college students were required to describe themselves for approximately a half an hour. Nonreinforcement, reinforcement of negative self-references, and reinforcement of positive self-references were compared. An interesting finding of this study is that although highly test-anxious subjects are extremely responsive to the reinforcement of negative self-references, they show no response to the reinforcement of positive self-references.

A large-scale investigation by Many and Many (1975) found a significant negative relationship between self-reported self-esteem and test-anxiety. Several studies provide even more direct support for the contention that high test-anxiety is associated with self-preoccupations and self-devaluing cognitions during task performance. Mandler and Watson (1966), Neale and Katahu (1968), and Marlett and Watson (1968) administered several tasks to high and low test-anxious individuals. In each investigation, posttask questionnaires asked subjects to indicate the incidence of self-devaluing cognitions during task performance. In all three studies, to the question, "How often during the testing did you find yourself thinking how well, or how badly you were doing?" the high test-anxious subjects reported a significantly higher incidence of such thoughts than did the low test-anxious persons.

Wine (1980) found that reassuring students improves their performance and that even relatively trivial manipulations, such as suggesting to students that they concentrate more on a task and less on extraneous matters, improves performance. Consistent with the framework promoted by this re-

searcher, such instructions enable students to devote a greater proportion of processing capacity to the task and less to anxious self-ruminations. The enhancement of performance may be attributable to having more cognitive capacity available for task solution.

The most direct evidence for the high incidence of interfering self-relevant thoughts during test-anxious persons' task performance was reported by Ganzer (1968). He recorded the task-irrelevant comments of high and low test-anxious individuals during a serial verbal learning task. The highly test-anxious subjects made many more irrelevant comments than did the low, and the comments were mostly of an apologetic self-deprecatory nature.

It is clear that not only is test anxiety associated with a general tendency to be self-deprecatory, but that negative self-devaluing cognitions are elicited in test-anxious individuals during evaluatively stressful cognitive task performance. It is suggested that the cognitive representation of these negative self-preoccupations must absorb some portion of the person's information processing capacity, thus leaving less capacity for coping with task demands. In this sense, anxious individuals must divide their attention between task requirements and irrelevant cognitive activities such as worry and self-criticism. With only partial attention devoted to the cognitive task, decrements in performance are anticipated, particularly on complex tasks that require complete attention for success. In this study the cognitive performance adequacy of high and low test-anxious subjects was measured and discussed in relation to levels of negative task-irrelevant thoughts.

Method

Subjects

The subjects were 214 students enrolled in an introductory psychology or sociology course at Grant MacEwan Community College. Three separate college campuses were involved in this study, with samples of students from south, central, and west Edmonton, Alberta. Grant MacEwan Community College offers one- and two-year programs in areas such as social work, law enforcement, and general arts for adults in postsecondary education. Twenty-one (nine males and 12 females) and 18 (nine males and nine females) volunteers were operationally defined as high and low test-anxious respectively on the basis of the Test Anxiety Scale (TAS). The subjects ranged in age from 18 to 46 years with a mean age of 27 years.

Procedure

The TAS (Sarason, 1973) was administered to the students, and the upper (scores>26) and lower (scores<9) 30% of the scale distribution operationally defined high and low test-anxiety groups respectively. The selection of score ranges for grouping subjects was influenced by several factors. First,

the score criteria and cut-off points that were used to define groups accounted for the anticipated attrition of subjects. Second, the mean anxieties of these groups are comparable to those of other studies by Sarason (1973) and Deffenbacher (1978). Finally, homogeneity in amount of test-anxiety experienced was established within low and high test-anxious groups (standard deviations of 2.3 and 3.1 respectively).

The 39 students, of which 21 were defined as high test-anxious and 18 defined as low test-anxious, were individually administered the Wisconsin Card Sorting Test (WCST) one to two weeks after the TAS was administered. It was initially explained to each participant that the author was conducting a research study designed to investigate individual reactions to tests. Students were told that the WCST was a measure of one's ability to do college-level work, and participants were instructed to perform to the best of their ability, as they would be compared with others. The WCST was presented in this way in an attempt to simulate an achievement-oriented situation.

The WCST provides six scores related to cognitive processing: inefficient initial conceptualization, number of perseverative responses, percent perseverative errors, percent conceptual level responses, failure to maintain

Table 1
Means and Standard Deviations for High and Low Test-Anxious Groups on WCST Subscales

Anxiety Level		WCST Subscales					
		1	2	3	4	5	6
High Anxious	$\bar{X}$	21.56	16.67	11.33	7.67	0.78	–1.33
	SD	17.41	9.11	1.71	1.49	0.81	9.00
Low Anxious	$\bar{X}$	32.62	27.19	35.71	6.29	2.14	–3.05
	SD	12.77	10.91	40.85	2.59	1.56	9.44
Effect Size		–0.73	–1.05	–1.15	0.68	1.14	0.19

WCST Subscales
1 = Number of perseverative responses
2 = Percent perseverative errors
3 = Inefficient Initial conceptualization
4 = Percent conceptual-level responses
5 = Failure to maintain set
6 = Learning across stages of the test

set, and inefficient learning across the stages of the test. The format of the test requires the subject to sort a pack of 64 cards on which are printed one to four symbols (star, cross, circle, or triangle), in green, red, yellow, or blue, under four stimulus cards (one red triangle, two green stars, three yellow crosses, and four blue circles) according to a principle that the subject must deduce from the pattern of the examiner's responses to the subject's placement of cards. For example, if the principle is color, the correct placement of a blue card is under four blue circles, regardless of the symbol or number. The examiner responds only with a "right" or "wrong" statement. After the subject has placed 10 consecutive cards according to the required principle, the examiner changes the principle being used for placement, indicating the change only in the changed pattern of the right and wrong statements given to the subject. The test continues until the subject has made six runs of 10 correct placements, has placed more than 64 cards in one category, or has used 128 cards.

Immediately after the WCST, the Cognitive Interference Questionnaire was administered. The CIQ (Sarason & Stoops, 1978) was designed to assess the degree to which people, after working on a task, report having had thoughts that interfere with concentration on the task. The CIQ consists of 22 items; the first 21 are rated for the frequency of occurrence of particular types of intrusive thoughts. The first 10 items pertain to the task on which the subject has just performed, and the remaining items refer to a diversity of thoughts unrelated to the task. The final item provides a global rating on a 7-point scale of the degree of mind-wandering experienced while working on a task. Testing time varied from 20 to 40 minutes per participant.

Two questions were asked in this study: (a) Does any difference exist between high and low test-anxious subjects' performance on the WCST? (b) Does a difference exist between the high and low anxiety groups as reported on the CIQ?

Results

Anxiety and Cognitive Performance

The Hotelling's T^2 was used to test the difference between groups on the six WCST scores, and the observed value of 25.46 was significant ($p<.007$). Overall, the test of the vector of means of the latent variable—cognitive processing—revealed significant differences between low test-anxious and high test-anxious groups. However, post-hoc tests failed to highlight any single variable as significant. Thus the six variables of the WCST as a group generate significant differences between high and low test-anxious groups, but no one variable is strong enough to explain the differences.

Level of Anxiety and Cognitive Interference

The TAS and the CIQ were used to measure level of test anxiety and amount of cognitive interference respectively. A comparison between the mean cognitive interference scores for the two anxiety groups indicated that high anxious individuals experienced a greater amount of cognitive interference than did their low anxious counterparts ($t=5.66$, df=37, $p<.0001$).

Discussion

The results of this study confirmed the anticipated performance decrements of highly test-anxious subjects but failed to identify a specific variable that was responsible. Also, as predicted, the high test-anxious group experienced greater interference from anxiety than did the low test-anxious group. Thus, for the highly test-anxious, an evaluative situation appeared to elicit interfering anxiety in the form of attention to worrisome thoughts and ruminations and focus on elements of the task irrelevant to efficient problem solution. The low test-anxious group, however, did not experience this dysfunctional pattern of attention.

The results of this study suggest a particular approach to the treatment of test-anxious persons, the goal of which is to reduce the negative self-focus, thus enabling students to devote a greater proportion of processing capacity to the task. The specific concern with how highly test-anxious subjects use their task time suggests treatment in which subjects are not only given intensive practice in dealing with tests, but also that this task practice should be accompanied by instructions to focus fully on the task and to inhibit self-relevant thinking. There is a large body of literature in selective attention, the basic premise of which is that the subjects' attention can be differentially directed to specific stimuli or away from others. With repeated training under attention-direction instructions, subjects become more and

Table 2

Cognitive Interference: Means and Standard Deviations for Low Anxious and High Anxious Groups

Anxiety Level	$\bar{X}$	SD	t
High Anxious	50.381	12.785	
			5.66*
Low Anxious	31.889	7.194	

*$p<.0001$

more skilled in attending to the "relevant" stimuli and ignoring the "irrelevant" stimuli (Wine, 1971). Further research is required to determine the degree to which attentional training not only improves task performance but lowers reported test anxiety levels as well.

Cognitive therapies implemented by school counselors or teachers offer effective coping strategies for test-anxious students. One such strategy involves replacing negative self-talk with positive self-talk, for example, "I am prepared; I have gone over this material thoroughly and will give this test my best shot." Cognitive therapy can also be used to counteract exaggerated irrational thoughts regarding possible negative test outcomes.

A weakness in the area of test anxiety research is that there has been no careful analysis of the nature of performance decrements experienced by test-anxious subjects. Focus has been placed merely on overall success or failure while ignoring particular sources of difficulty. The WCST, which provides scores for six variables, was used in this study in an attempt to overcome this weakness. Future researchers need to recognize and define the properties of specific tasks and the attentional demands they make on subjects.

References

Deffenbacher, J.L. (1978). Worry, emotionality, and task-generated interference in test anxiety: An empirical test of attentional theory. *Journal of Educational Psychology, 70*(2), 248-254.

Ganzer, V.J. (1968). *The process of test anxiety: An investigation*. New York: Longfellow-Verlageist.

Mandler, G., & Watson, D.L. (1966). Anxiety and the interruption of behavior. In C.D. Spielberger (Ed.), *Anxiety and behavior* (pp. 263-288). New York: Academic Press.

Many, M.A., & Many, W.A. (1975). The relationship between self-esteem and anxiety in grades four through eight. *Educational and Psychological Measurement, 35*, 1017-1021.

Marlett, N.J., & Watson, D. (1968). Test anxiety and immediate or delayed feedback in a test-like avoidance task. *Journal of Personality and Social Psychology, 8*, 200-203.

Neale, I.M., & Katahu, M. (1968). Anxiety, choice and stimulus uncertainty. *Journal of Personality, 36*, 235-245.

Sarason, I.G. (1960). Empirical findings and theoretical problems in the use of anxiety scales. *Psychological Bulletin, 57*, 403-415.

Sarason, I.G. (1973). The test anxiety scale: Concept and research. In I.G. Sarason & C.D. Spielberger (Eds.), *Stress and anxiety* (Vol. 5, pp. 193-216). Washington, DC: Hemisphere.

Sarason, I.G., & Stoops, R. (1978). Test anxiety and the passage of time. *Journal of Consulting and Clinical Psychology, 46*(1), 102-109.

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin, 76*(2), 92-104.

Wine, J. (1980). Cognitive-attentional theory of test-anxiety. In I.G. Sarason (Ed.), *Test anxiety: Theory, research applications* (pp. 349-385). Hillsdale, NJ: Erlbaum.

The Alberta Journal of Educational Research *Volume XXXV, No. 2, June 1989, 171-186*

LESLIE D. VAALA

and

EDWARD A. HOLDAWAY
University of Alberta

The College to University Experience: Satisfaction and Success of Students Who Transfer

College transfer students attending the University of Alberta were surveyed to determine which variables were most closely associated with student success and satisfaction in their transfer from college to university. Both interviews and questionnaires were used. Students tended to be more satisfied with the college experience than with the university experience. Most students reported using a different form of support system at university than at college. Students who were more successful at university had a greater preference for a transfer program than did the less successful students. Implications for practice are related to availability of program counseling and transfer information. Suggestions for further research include the development of a student decision making model.

In varying degrees, universities in North America have attracted students who begin their studies in a college setting and, after a period of study, move to a degree-granting university to complete their programs. These students and their programs are a significant presence in contemporary postsecondary education. The purpose of the present study was to examine some of the factors that influence the satisfaction and success experienced by students who attend college before university.

For several decades, transfer education has received wide attention in the literature. Dennison (1986) stated that Canadian college enrollments have grown continuously in recent decades and noted that the transfer function of British Columbia colleges has been challenged by the addition of other college programs. He observed, however, that although colleges offer increasingly diverse programs to larger numbers of students, the number of British Columbia college students annually registered in transfer programs during this period has remained in the 7,500 to 9,000 range.

On the American scene, Cohen (1985, p. 152) found that college enrollments have grown continuously during the past four decades. For this period, Knoell (1982) concluded that the transfer function of American colleges has been diminished by the expansion of other college programs. The

---

Leslie Vaala, whose interests are in student services and postsecondary administration, is a sessional lecturer in educational administration at the University of Alberta.

Edward Holdaway is a professor of educational administration at the University of Alberta and has interests in satisfaction of educators, administration of postsecondary education, and organizational analysis.

change in college program offerings appears to reflect changes in the types of expectations that have been placed on colleges. One consequence of the greater demand being placed on the college system in Canada and the United States may be that transfer students are not receiving as much institutional support as they once did.

Research has suggested that transfer students may experience special problems in completing their educational programs. Dennison and Jones (1970) identified lack of academic preparedness, loss of course credits, and lack of prerequisites as problems affecting British Columbia students in the 1960s. Fetters (1977), in a national longitudinal study in the United States, found that two-year transfer-in students withdrew more frequently from a four-year campus than did students who began their study on a four-year campus. Cross (1968), Cohen and Brawer (1981), and Rich (1979) concluded that ethnicity, finances, residence location, and admissions policies also contributed to the particular problems experienced by transfer students.

The purpose of this article is to report a recent study of transfer students attending the University of Alberta and to discuss implications of findings about transfer student satisfaction and success for the postsecondary system in Alberta.

Relevant Literature

Studies of the transfer function have progressed over time from initially investigating issues involving transfer of credits (Cross, 1968; Knoell & Medsker, 1965) to measuring completion rates in baccalaureate degree programs (Dennison, 1974) to present research interests of identifying problems experienced by transfer students in the transfer process (Remley & Stripling, 1983). Literature related to university transfer students has been summarized in the areas of (a) transfer in Alberta and (b) satisfaction and success in university.

Transfer in Alberta

In a critical analysis of the Canadian college system, Dennison (1986) observed that there is no national college-level system of education for Canada. Rather, he noted that the role for colleges differed in each province, with the transfer function strongest in Alberta, British Columbia, and Quebec.

In September 1974, at the request of the Universities Coordinating Council and presidents of institutions of higher education in Alberta, the Minister of Advanced Education announced the establishment of a Council on Admissions and Transfer (Alberta Council on Admissions and Transfer, 1981). From the several pressing transfer-related issues, the Council decided that college-university transfer problems were a first priority. This priority was addressed by preparing and publishing a *Provincial Transfer Guide* for the 1976-1977 academic year. The guide, now entitled *Alberta Transfer Guide*, is published annually to reflect the current status of negotiated arrangements between colleges and universities in the province (Alberta Council on Admissions and Transfer, 1987).

A transfer patterns study conducted in 1983 provides the most recent figures about the volume, direction, and nature of postsecondary student transfer in Alberta. The findings of this study (Alberta Advanced Education, 1984) revealed that the largest proportion of transfers over a five-year period had been to universities. Of the 5,136 students who transferred in the postsecondary system in 1979, 4,144 students transferred to universities. By 1983, out of a systemwide total of 7,078 student transfers, the number of university transfers had grown to 6,987, an increase of 69% compared to a systemwide increase of 38%. According to the Council, "the single institution receiving the highest number of full-time transfers each year was the University of Alberta" (Alberta Advanced Education, 1984, p. 22).

After a three-year study of students transferring from an Edmonton community college to the University of Alberta, Small and Konrad (1986) concluded that there were sufficient issues about student persistence and satisfaction to warrant further study.

Satisfaction and Success in University

The study of student satisfaction and performance is related to, but distinct from, worker satisfaction and performance. Tinto (1986) succinctly described this difference and cautioned against direct application of job satisfaction research to student satisfaction research by stating that

> The application of theories of work organizations ... must be carried out with care. The primary difficulty with such applications is that they make the implicit assumption that higher education organizations are essentially the same as those in the world of work and, therefore, that one can think of students in those organizations as one would of workers in factories or offices. While the analogy might be stretched to fit faculty and staff, it is doubtful that students would see themselves in the same light as would workers generally. Though the analogy of worker productivity to student performance is especially appealing ... we must be careful not to push such analogies too far. (p. 377)

Student satisfaction research has not been explicitly predicated on the models of worker satisfaction. Most studies of student satisfaction are descriptive research. One of the most thorough analyses of accumulated research about students in American colleges and universities was conducted by Feldman and Newcomb (1969) in the late 1960s. After reflective consideration of what was known at that time about the impact of higher education of students, they wrote that

> Our best guess at the moment is that a college is most likely to have the largest impact on students who experience a continuing series of not-too-threatening discontinuities. Too great a divergence between student and college, at least initially, may result in the marshalling of resistances. Too little might mean no impetus for change. From this point of view, a college's objectives might include a tolerance, or even a desire, for those discrepancies that can stimulate change and growth. (p. 332)

Studies of student satisfaction have continued in this best-guess-at-the-moment approach. In these descriptive studies, satisfaction has been defined in various ways. For example, Aitken (1982) and Babbitt and Burbach (1985) take the term to mean student acceptance of academic

programs and living conditions. Bean and Metzner (1985, p. 523) have described satisfaction as "the degree to which a student enjoys the role of being a student and reports a lack of boredom with college courses." A further example is provided by Nafziger, Holland, and Gottfredson (1975, p. 132) who defined a satisfied student as one who is a "typical student at his college and [has] a personality pattern which is both consistent and well defined."

Astin (1974, 1978), after examining overall student satisfaction in a national sample of university students from 194 American institutions, concluded that students' overall satisfaction with university could be measured. Astin also identified specific facets of satisfaction that could be measured including teaching, curriculum, facilities, career preparation, extracurricular activities, and administrative services. He found that little variance in student satisfaction could be accounted for by entering characteristics, institutional characteristics, and measures of student involvement (1974, p. 168). Specific aspects of the college experience (ranked in Astin's findings) which did have substantial effects on satisfaction were (a) quality of instruction; (b) contacts with faculty and fellow students (Pascarella, 1980, has suggested that faculty and peer contacts have the greatest positive effects that add significantly to overall satisfaction); (c) curriculum; (d) college administration; (e) reputation of the institution (this effect is greatest among students with higher grades); and (f) intellectual environment.

Willingham (1985) reported on a nine-institution longitudinal study related to success in university. These universities and colleges, with the exception of one California college, were all located in the northeastern United States. The purpose of the study was to identify precollege characteristics that were reliable predictors of success in undergraduate programs. He found that each institution had individualized definitions of student success. Generally, three types of success were recognized: academic, leadership, and significant personal accomplishment. Characteristics of successful students were determined from studying a group of students identified by faculty, administrators, and other students as the most successful students on each of the nine campuses. That study found that the precollege characteristics which correlated most highly with a student's subsequent identification as successful by college faculty at the end of an undergraduate program were (a) high school rank, (b) SAT scores, (c) high school honors, and (d) indications of follow-through in work or activities.

Research Project

A study of college transfer students attending the University of Alberta was conducted during the 1987-88 winter session (Vaala, 1988). In this study, "transfer program" was defined as a university-level program of study that commences with enrollment at a college and culminates at a university where the remaining courses required for a specific degree are completed. "Transfer experience" was defined as a period of transition that begins when a college student initiates application to the university and ends when that student personally identifies as a university student. "College" was

defined as a nonuniversity, postsecondary educational institution offering university-recognized, university-level credit courses.

Purpose and Design

The purpose of this study was to use student perceptions to identify, describe, and categorize those variables which are most closely associated with success and satisfaction in transfer from college study to a university-based degree program. Transfer students' perceptions were separated into background factors and outcome factors in order to examine possible relationships between perceptions of satisfaction and measures of success in the transfer program experience.

A conceptual framework of the major theoretical elements of the study is presented in Figure 1. Relationships were assumed to exist between selected student characteristics (e.g., educational goals, career goals, student housing, work experiences, previous academic performance, and individual attributes) and aspects of students' transfer programs and transfer experiences. A relationship also was assumed to exist between selected supports (e.g., family, peers, and staff) and the transfer program. A further relationship was assumed to exist between a transfer program and transition from college to university as moving from one campus to another is inherent in the concept of a transfer program. Perceptions of satisfaction and success in the transfer program and experience were seen as self-determined criteria of students. While a relationship between satisfaction and success was assumed, this was not viewed as causal. Satisfaction and success were seen to be overall outcomes of both the transition and the transfer program.

Participants

From a cohort of 1,081 transfer students at the University of Alberta, two random samples were drawn: (a) 400 students received questionnaires in October 1987, and (b) 100 students were invited to participate in an interview. The response rate for usable questionnaire returns was 67.3%. To clarify and enrich interpretations from the questionnaire data, interviews were conducted in January 1988 with all 10 students who volunteered to be interviewed.

The ratio of female to male respondents to the questionnaire was higher than the ratio between the sexes in the population for this study (3:2 versus 1:1). The average age of these students was 23.5, with a range from 17 to 47. A large majority (82.9%) of the respondents was single. The mean high school matriculation average of these students was 70.9%. Although students transferred to university from a total of 14 Alberta colleges (four Edmonton colleges and 10 colleges throughout other areas of Alberta), over 60% of the respondents transferred from one non-Edmonton and two Edmonton colleges.

The interview participants resembled the questionnaire sample with respect to sex, age, degree program, type of college attended, and location of college attended.

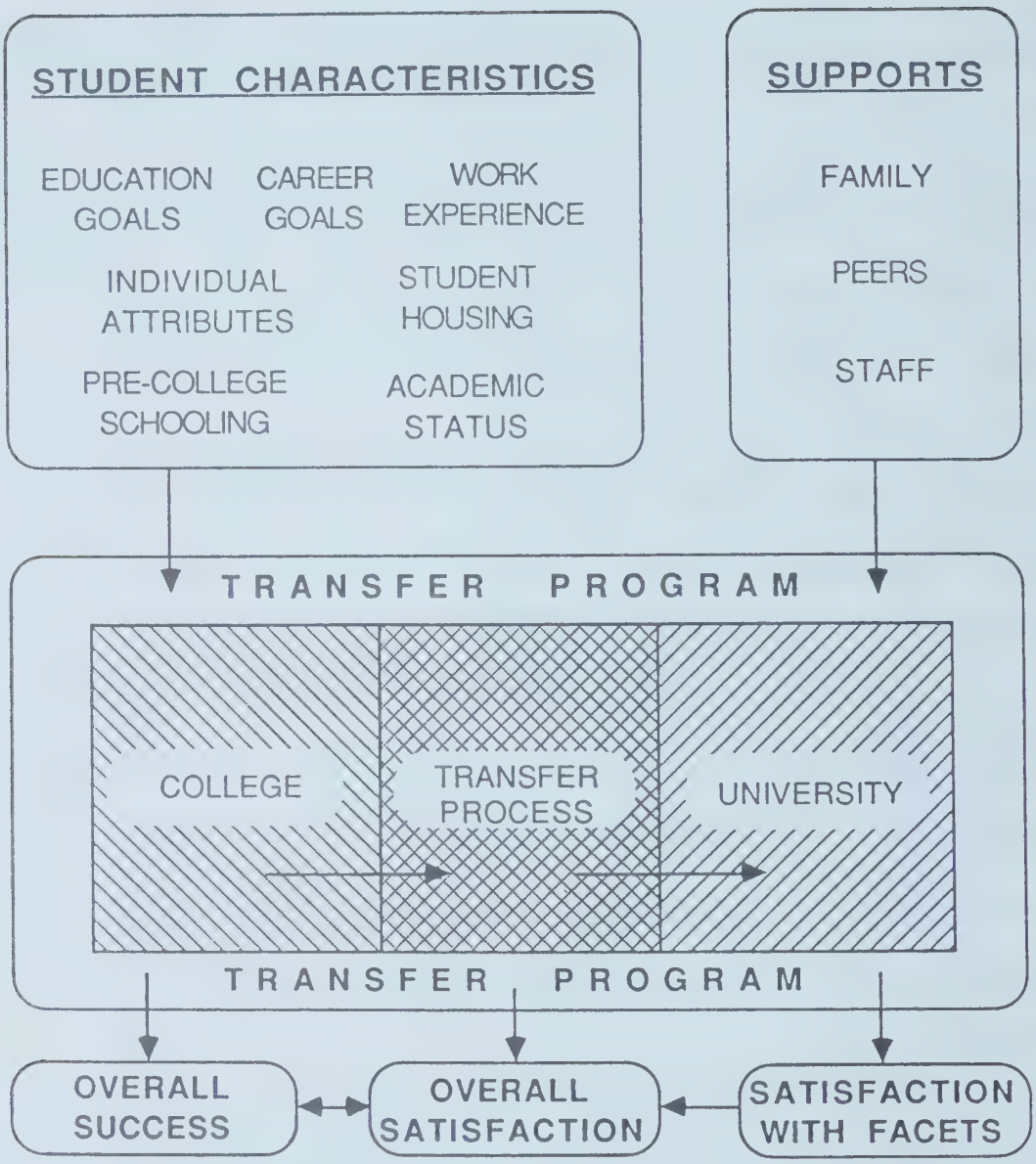


Figure 1: Conceptual framework of the relationships among variables involved in the transfer process

Data Collection and Analysis

A 64-item questionnaire was developed for the study. Demographic data and information about attendance patterns were collected with 36 items requiring a choice from a list or a short written response. Data about goals and satisfactions were collected by 18 items with five-point Likert-type scales. There were six items about sources of assistance and goals also requiring participants to choose a response or write a short response. Finally, four items about transfer problems and preferences invited participants to provide longer written responses.

An interview schedule addressing five broad areas was developed following initial analysis of questionnaire data. Two areas focused on experiences and definitions of satisfaction and success. Two areas were about experien-

ces in adjusting to university after transfer. The fifth area was about use of supports and assistance during transfer.

In addition to examining the results for the total sample, several groupings of respondents were developed as a means of exploring relationships between respondent characteristics and major variables in the study. These groups were formed on the basis of attendance patterns, college location, college type, degree program, and date of first university registration.

Findings

The major findings are summarized in five parts: (a) satisfaction with the transfer experience, (b) satisfaction with the transfer program, (c) success with the transfer experience, (d) success in the transfer program, and (e) use of supports and assistance by students in transferring.

Satisfaction with the transfer experience. A majority of respondents (58%) reported that they were satisfied with the overall process of transferring to university. Students who attended university transfer colleges and university/technical colleges (i.e., colleges offering only university-level courses and colleges offering university-level as well as vocational/technical courses) tended to report greater satisfaction with the transfer experience than did those who attended technical colleges (i.e., colleges offering only vocational/technical courses).

Students tended to define satisfaction in the transfer experience as receiving university credit for courses taken at college and reaching the stage of feeling comfortable at the university. A large majority (83.7%) of students reported some difficulties in completing the transfer process, with the most frequently cited being adjusting at university to (a) larger class sizes, (b) less personal professor-student interactions, and (c) heavier academic workloads.

Satisfaction with the transfer program. Students most satisfied with the transfer program were those who attended a university/technical college for two or four winter terms, transferred at least 27 credits to university, and then attended university continuously for at least three full winter terms.

A comparison of student perceptions of satisfaction at college and university with five aspects of the transfer program is presented in Table 1. The results of a correlated *t*-test reveal a statistically significant greater satisfaction was expressed with four of these items at college than with these same items at university. Students tended to report higher satisfactions from interactions with faculty and peers at college than with similar interactions at university. Students also tended to report more satisfaction with their intellectual development at college than at university. Positive faculty influences on students were perceived to be greater at college than at university. Generally, students were satisfied with decisions to attend both college and university.

Further, satisfactions with facets of the transfer program correlated more highly at college than at university. Figures 2 and 3 present Pearson correlation coefficients among comparable aspects of the college and university experiences. With respect to both college and university, the highest

Table 1
Means of Respondents' Perceptions of Satisfaction with Five Aspects of the Transfer Program at College and University

Aspects	Location of student experiences	
	College	University
The student friendships I developed have been personally satisfying	4.2	3.73**
I am satisfied with the extent of my intellectual development since enrolling	4.03	3.86*
Faculty members with whom I have had contact are genuinely interested in students	4.13	3.31**
My interactions with faculty members have had a positive influence on my career aspirations	3.70	3.34**
I am confident that I made the right decision in choosing to attend	4.23	4.32

Note: Five-point scale, 1 = strongly disagree, 5 = strongly agree
*p < .05, **p < .01

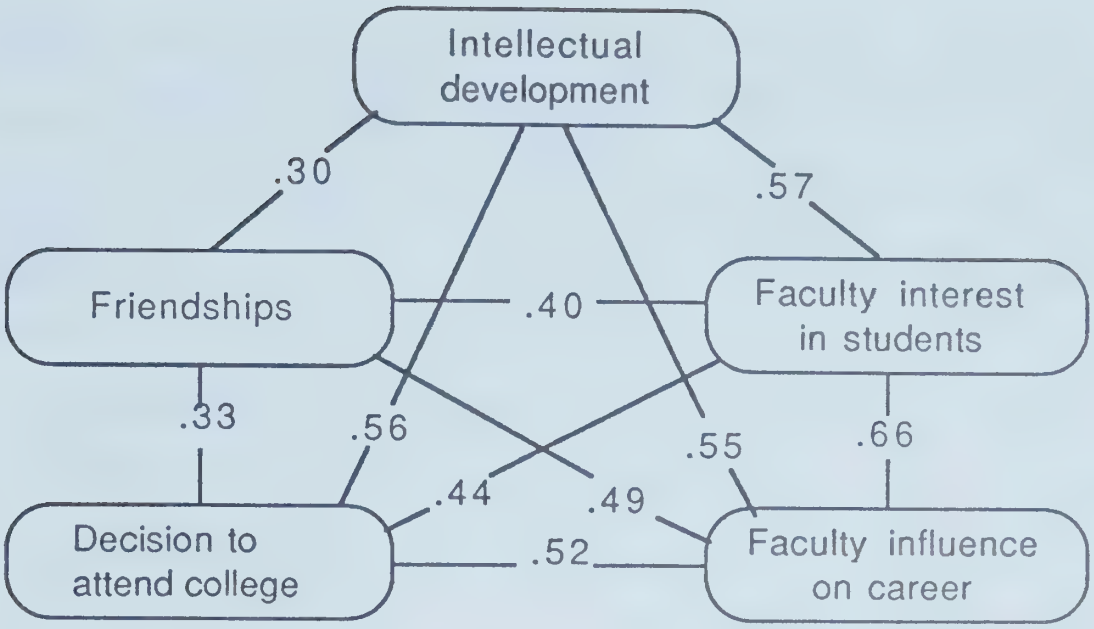
correlations of responses tended to involve pairs which included one of the two items about faculty members.

Overall, students tended to report high satisfaction with the college experience. Only 26% of the students indicated that their present preference was to have taken all courses at the university. The remaining majority reported a preference for having attended college before university.

Success with the transfer experience. Most students perceived that they were able to obtain adequate information about transfer while at college. Students most likely to hold this perception had attended a university transfer college or university/technical college full time for two or four winter session terms and transferred to university within a year of leaving college.

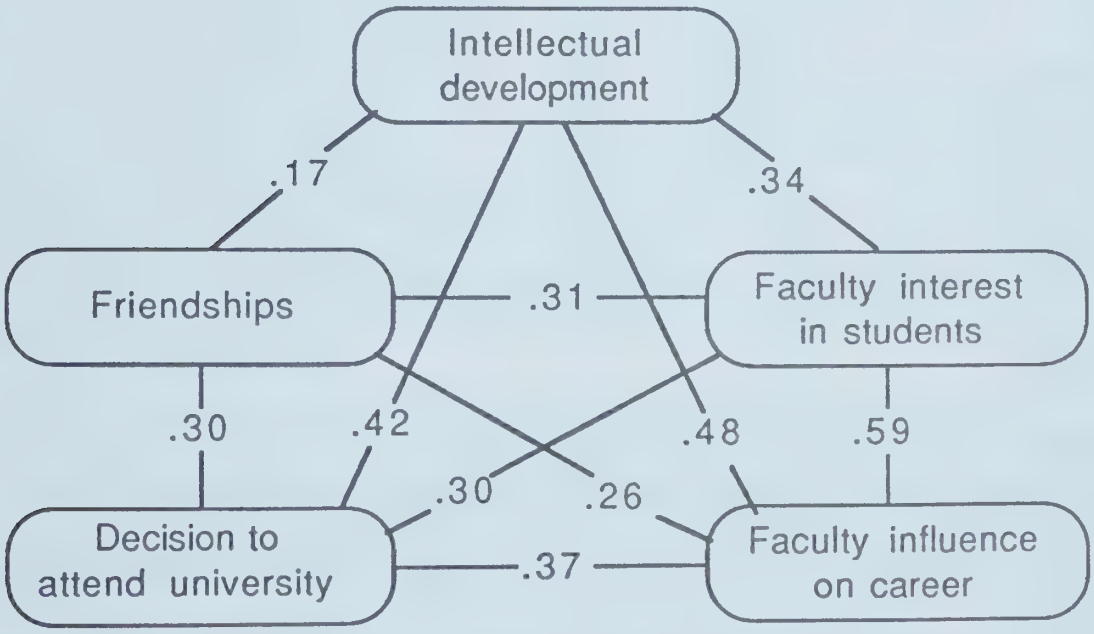
About half the participants visited the university before transferring. These students tended to report such visits as beneficial to their transfer experience. Also, before transferring, most students had some contact with persons at the university; friends were the most frequent contact. Technical college students were the least likely and university transfer college students were the most likely to have friends already registered at university.

Most students reported that they felt at ease and had developed an identity as a university student by the end of their first winter term at university. Students tended to perceive the major adjustments to university as including (a) larger classes, (b) heavier workloads, and (c) less personal professor-student interactions.



Correlation coefficients of .126 and .168 were needed for statistical significance at the .05 and .01 levels

Figure 2: Pearson correlation coefficients among satisfaction with facets of the transfer program at college



Correlation coefficients of .126 and .168 were needed for statistical significance at the .05 and .01 levels

Figure 3: Pearson correlation coefficients among satisfaction with facets of the transfer program at university

Success in the transfer program. Respondents tended to report a decline in grades from college to university, with the mean decline in grade point average being 0.5 on a 9-point grading scale. This overall decline appeared unrelated to sex, degree program, college attended, or attendance patterns.

Students who attended university transfer colleges or who attended college full time for two or four winter-session terms tended to transfer the equivalent of one or two years of study toward their university degree. These students were also most likely to carry and complete a full-course load (i.e., 15 credits per term for three winter session terms).

Using the measures of credits completed and grade point average, students who were more successful in the transfer program tended to (a) report higher personal educational goals, (b) perceive that their earnings went to pay their educational costs, (c) report a greater preference for a transfer program, and (d) have parents who were more likely to have completed a postsecondary program.

Use of supports and assistance. Students tended to use sources of assistance differently at university than at college. They frequently mentioned the positive value of college faculty influences, particularly in program and career planning. At college, friendships with peers were generally a source of social and academic support. College registrars were described as more helpful than college counselors in providing information about transfer programs and transferring.

Difficulties in the transition from college to university were reported by 83.7% of the students (Table 2). Over half these students reported more than one difficulty. With few exceptions, all difficulties were related to situations encountered in Edmonton or at university. Nearly a quarter of the reported difficulties (21.6%) related to the personal adjustments of leaving one social environment and entering another. The largest category of difficulties (30.7%) related to the academic aspects of the transfer program. Slightly less than a quarter of reported difficulties (23.1%) related to the physical surroundings of the university, while almost a quarter (24.6%) were about administrative services or procedures related to the transfer process.

Over half the students commented that increased college and university attention could be given to providing students with current and accurate information related to transferring. By contrast a minority of students indicated that the college and the university did not need to provide new or additional assistance for students who transfer (34.2% and 21.2% respectively).

Most students described using a different form of support system at university than at college. University faculty members were perceived less frequently than college faculty members as providing assistance in program or career planning. University administrative faculty members (e.g., assistant deans) tended to be perceived as more helpful than were administrative staff (e.g., receptionists and clerks). Peer contacts at university were generally described as useful for sharing academic information; social friendships were frequently a continuation of friendships formed at college.

Table 2
Frequency of Difficulties Faced in Transferring to University
(n = 216)

Category Difficulty	f	By category %f	Overall %f
Personal			(21.6)
Forming new friendships	25	24.5	5.3
Moving to a new city	17	16.7	3.6
Financing educational costs	17	16.7	3.6
Loss of contact with family/friends	14	13.7	3.0
Finding accommodation	13	12.7	2.8
Arranging transportation	5	4.9	1.1
Leaving employment	4	3.9	0.8
Other	7	6.8	1.4
Academic			(30.7)
Larger class size at university	39	26.9	8.3
Heavier workload requirements	28	19.3	5.9
Lack of interaction with professors	21	14.5	4.4
Greater academic competition	17	11.7	3.6
Readjustment to academic life	15	10.3	3.2
Adjustment to a new grading system	10	6.9	2.1
Writing Competency Test requirement	6	4.1	1.3
Adjustment to new teaching methods	5	3.4	1.1
TOFEL requirements	4	2.8	0.8
Campus environment			(23.1)
Impersonal nature of the campus	38	34.9	8.1
Overwhelming campus size	33	30.3	7.0
Unfamiliarity with facility locations	27	24.8	5.7
Unfamiliarity with locating materials	7	6.4	1.4
Rapid pace of activity/events	3	2.8	0.6
Parking	1	0.9	0.2
Administration			(24.6)
Lack of assistance to plan program	33	28.4	7.0
Loss of course credits	29	25.0	6.1
Late official acceptance	14	12.1	3.0
"Red tape"	12	10.3	2.7
Requirements for applications	9	7.8	2.0
Registration procedures	9	7.8	2.0
Other	10	8.6	2.1
Total	485		100

Note: Some students provided more than one response.

Family members, choice of residence, and financial assistance were seldom mentioned as influential forces in program decisions or program completion.

Implications

The transfer student study has several implications for practice as well as for theory and further research.

Context for the Future

The future direction of postsecondary education in Alberta rests in part on the increased availability of publicly accepted transfer program opportunities. Long-range planning at the University of Alberta, the second largest university in Canada, projects a staged decrease in undergraduate enrollments in order to accommodate an increased institutional emphasis on graduate programs (University of Alberta, 1987). Similarly the University of Calgary, a large and growing Canadian university, has been gaining increased national and international recognition as a research-oriented university. As these major universities move to limit undergraduate enrollments, other institutions in the Alberta postsecondary system are expanding their undergraduate programs.

Establishment of the Alberta Private Colleges Accreditation Board in 1983 (Government of Alberta, 1984) has enabled three private junior colleges to attain four-year degree-granting status. In their role as colleges affiliated with the University of Alberta, these colleges previously admitted substantial numbers of transfer students. For selected degrees, students are now able to complete all undergraduate work at these colleges, which would serve potentially to lower the number of transfer students moving to universities. However, the new degree-granting status of these colleges may also serve to expand further public awareness and acceptance of Alberta colleges in general as "stepping stones" to university study.

Further, despite provincially imposed funding restrictions on the expansion of programs at Alberta public colleges, an Edmonton community college was authorized in early 1988 by the Department of Advanced Education to establish university transfer courses for several hundred transfer students, effective September 1988. While this move may also have been intended to alleviate concerns of overcrowding at the University of Alberta, it too may contribute to further enhancement of the public's acceptance of transfer programs. Thus, in an era when the transfer function of colleges is being actively questioned in the literature (Cohen, 1985; Kissler, 1982; Knoell, 1982), postsecondary education policy in Alberta appears to be strengthening rather than diminishing transfer opportunities for students. Consequently, colleges will have to adjust to the systemwide trend to accept more postsecondary students in transfer programs.

There are implications from this study for two broad areas of transfer programs: faculty-student interactions and transfer-program information sources.

Faculty-student interactions. Faculty members appear to contribute considerably to student decision making about program and career choices,

particularly at the college level. This activity should continue to receive institutional recognition and support. The positive outcomes of faculty-student interactions with respect to program advising and career planning in the college setting should be widely acknowledged and encouraged. Supports and assistance should be routinely made available to individual faculty members to ensure than they have current knowledge about transfer programs. Further, university faculty members may need to conspicuously identify ways and means of ensuring that transfer students are accorded opportunities equal to those of other students with respect to interacting with individual faculty members about concerns over program and course information.

Transfer program information. Transfer students have identified the positive value of visiting the university campus before transfer and of making contact with administrative faculty in appropriate academic departments. In many instances, individual students have initiated these activities; in other cases, college staff have coordinated these events. These present forms of contact should be strengthened and encouragement provided for college and university staff to develop increased opportunities for transfer students to visit the university campus before commencing courses at that location.

Although many students were found to be aware of the *Alberta Transfer Guide* (Alberta Council on Admissions and Transfer, 1987), they held diverse interpretations about the guide's purpose. To resolve many student concerns related to transfer of credit, copies of the guide should be made available at college to individual students. Some difficulties in transferring have arisen when students were not given direct access to the guide. In addition, the guide should clearly and conspicuously inform its readers about how it is to be used. For example, some difficulties have arisen when students generalize from the acceptance of a specific college course for a specific university degree program to a conclusion that the course is also acceptable for other university degree programs.

Implications for Theory and Research

Further transfer student study should focus on the process of student decision making with respect to transfer program decisions. The finding that many students viewed a college experience as a preparatory step for university prompts a call for research to build on Zemsky and Oedel's (1983) model of the structuring of college choice in order to address student anticipations/expectations when deciding to attend college before university and when deciding which university to attend. Zemsky and Oedel's perspective of the elements that influence college choice may identify important dimensions to include in a student decision making model. Additional study at this juncture could add valuable insights to the literature about college choice in areas such as student-institution match and institutional response to changing student demographics.

Investigation also needs to move beyond examination of the functional outcomes of faculty advising (Hornbuckle, Mahoney, & Borgard, 1979;

Stickle, 1982) by developing credible perspectives on the mentoring/modeling role of faculty in colleges with respect to student career and program decisions. (For example, what are the activity patterns of faculty with respect to assisting transfer students? Are faculty equally accessible to all transfer students? What factors are associated with effectiveness?) Volkwein, King, and Terenzini (1986) have begun investigation of transfer student intellectual development by focusing on student-faculty interactions before and after transfer. In the same light, expanded attention should be given to identify the dimensions of college and university faculty influences on transfer students. For example, what is the scope, nature, and intensity of this influence with respect to choice of courses, choice of university, and the shaping of expectations about the nature of the university portion of the experience?

Formulation of a transfer student decision making model also should incorporate the nonacademic aspects of the college-university experience which Pascarella (1986) suggests students perceive as substantial influences on their decision making processes. For example, this study identified some variables, such as parental completion of postsecondary education programs, aspirations for degrees beyond the bachelor's level, and students' intention/perception of paying their own education costs, that may be associated with students who were more successful in the transfer program. Similarly, Willingham (1985) concluded that student success contained academic, leadership, and personal accomplishment elements. Although participants in this study seldom reported on the development of leadership skills, greater attention should be given to examine the influence of personal accomplishments and leadership opportunities on transfer student decision making.

Concluding Comments

A study of transfer students assessed the effectiveness of several dimensions of Alberta college-based university programs. The substantial positive influence of college faculty members on transfer students is a reflection of the support and encouragement available to students within the college system. Lack of student recognition for other support services at colleges may suggest that these are taken for granted or not extensively used by transfer students.

The effects of college socialization appear to have strongly influenced transfer students. For example, students tended to strongly affirm that the college contributed positively to their intellectual development. Compared to college, however, the university portion of the transfer program experience produced a disproportionately large number of negative comments from transfer students. Although students "moved on" because they felt they were ready for university, they also appeared to find university substantially different from what they had anticipated.

This perspective openly questions the notion that transfer students embrace the university portion of their studies. In the Alberta postsecondary system which appears to be placing greater emphasis on the transfer function of colleges, university administrators need to pay close attention to cir-

cumstances surrounding transfer student satisfactions in order to ensure that enrollment pattern projections are actually achieved. Clearly, there is a need to involve more institutions and to continue a more detailed examination of variables relevant to the transfer experience and of the students who participate in this experience.

References

Aitken, N.D. (1982). College student performance, satisfaction, and retention: Specification and estimation of a structural model. *Journal of Higher Education, 53*(1), 32-50.

Alberta Advanced Education. (1984). *Participation patterns study: Report of the committee to examine participation trends of Alberta post-secondary students*. Edmonton: Author.

Alberta Council on Admissions and Transfer. (1981). *Transfer patterns: A study of student transfer among Alberta post-secondary institutions*. Edmonton: Alberta Advanced Education.

Alberta Council on Admissions and Transfer. (1987). *Alberta transfer guide, 1987-1988: For post-secondary institutions*. Edmonton: Alberta Advanced Education.

Astin, A.W. (1974). Measuring the outcomes of higher education. In H. Bowen (Ed.), *New directions for institutional research: Evaluating institutions for accountability* (pp. 23-46). San Francisco: Jossey-Bass.

Astin, A.W. (1978). *Four critical years: Effects of college on beliefs, attitudes, and knowledge*. San Francisco: Jossey-Bass.

Babbitt, C.E., & Burbach, H.J. (1985). Women and the college environment: A study of organizational satisfaction, compliance, and academic achievement. *Sociological Focus, 18*(1), 37-47.

Bean, J., & Metzner, B.S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*(4), 485-540.

Cohen, A.M. (1985). Student access and the collegiate function of community colleges. *Higher Education, 14*(2), 149-163.

Cohen, A.M., & Brawer, F.B. (1981). The persistent issues. *Community and Junior College Journal, 54*(4), 17-21.

Cross, K.P. (1968). *The junior college student: A research description*. Princeton, NJ: Educational Testing Service.

Dennison, J.D. (1974). Characteristics of college students. *Yearbook of the Association of Community Colleges*. Edmonton: University of Alberta.

Dennison, J.D. (1986). *Canada's community colleges: A critical analysis*. Vancouver: University of British Columbia Press.

Dennison, J.D., & Jones, G. (1970). *The community college transfer student at the University of British Columbia: A three year study*. Vancouver: Vancouver City College.

Feldman, K.A., & Newcomb, T.M. (1969). *The impact of college on students* (Vol. 1). San Francisco: Jossey-Bass.

Fetters, W.B. (1977). *Withdrawal from institutions of higher education: National longitudinal study*. Washington, DC: U.S. Department of Health, Education and Welfare.

Government of Alberta. (1984). *Universities act*. Edmonton: Queen's Printer.

Hornbuckle, P.A., Mahoney, J., & Borgard, J.H. (1979). A structural analysis of student perceptions of faculty advising. *Journal of College Personnel, 20*(4), 296-300.

Kissler, G.R. (1982). The decline of the transfer function: Threats or challenges? *New Directions for Community Colleges, 10*(3), 19-29.

Knoell, D.M. (1982). The transfer function—one of many. *New Directions for Community Colleges, 10*(3), 5-17.

Knoell, D.M., & Medsker, L.L. (1965). *From junior college to senior college: A national study of the transfer student*. Washington, DC: American Council on Education.

Nafziger, D.H., Holland, J.L., & Gottfredson, G.D. (1975). Student-college congruency as a predictor of satisfaction. *Journal of Counseling Psychology, 22*(2), 132-139.

Pascarella, E.T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research, 50*(4), 545-595.

Pascarella, E.T. (1986). College environmental influences on learning and cognitive development: A critical review and synthesis. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 1, pp. 1-62). New York: Agathon.

Remley, T.P., & Stripling, R.O. (1983). Perceptions of transfer problems experienced by community college graduates. *Journal of College Student Personnel, 24*(1), 43-49.

Rich, I.A. (1979). Counseling the transfer student: Myths and realities. *Journal of College Student Personnel, 20*(2), 175-176.

Small, J.M., & Konrad, A.G. (1986). *Transfer study: Final report*. Edmonton: University of Alberta.

Stickle, F. (1982). Faculty and student perception of faculty advising effectiveness. *Journal of College Student Personnel, 23*(4), 262-265.

Tinto, V. (1986). Theories of student departure revisited. In J.C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 2, pp. 359-384). New York: Agathon.

University of Alberta. (1987). *The next decade and beyond: Draft policies*. Edmonton: Author.

Vaala, L.D. (1988). *Satisfaction and success of university students in transfer programs*. Unpublished doctoral dissertation, University of Alberta, Edmonton.

Volkwein, J.C., King, M.C., & Terenzini, P.T. (1986). Student-faculty relationships and intellectual growth among transfer students. *Journal of Higher Education, 57*(4), 413-430.

Willingham, W.W. (1985). *Success in college*. New York: College Entrance Examination Board.

Zemsky, R., & Oedel, P. (1983). *The structure of college choice*. New York: College Entrance Examination Board.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

- Manuscript in triplicate.
- Typewritten, double-spaced throughout.
- Abstract of approximately 100 words, on a separate sheet.
- Manuscript not to exceed 6,500 words excluding graphics.
- Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

- Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s).
- Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format.
- All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
- Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes*; any citations in notes follow the same format as other references.

Graphics

- Each table and figure on a separate page at the end of the manuscript.
- Figures, charts, and diagrams in camera-ready form.
- Number tables in Arabic numerals and center *Table 1* over each heading.
- Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1*: ...
- Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

- *AJER*'s editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
- Editorial changes may be made to manuscripts.
- For spelling, consult *Webster's New Collegiate Dictionary*; spelling in quoted material remains as in the original.

Diskettes

- If manuscript is produced on an IBM, IBM compatible, or Macintosh computer, please also send diskette. Microsoft Word, MultiMate, WordPerfect, WordStar, Xerox Writer, XyWrite, and DisplayWrite 3 are acceptable.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*